Spring 5-22-2015

A Geospatial Analysis of Norovirus Outbreaks in California, and an Investigation of the Impact of Environmental Variables

Alexandra Groen
University of San Francisco, algroen@usfca.edu

This Master's Project

A Geospatial Analysis of Norovirus Outbreaks in California, and an Investigation of the Impact of Environmental Variables

by

Alexandra Groen

is submitted in partial fulfillment of the requirements
for the degree of:

Master of Science
in
Environmental Management

at the

University of San Francisco

Submitted:	Received:
.....	.....
Alexandra Groen May 21, 2015	Cindy Lowney, Ph.D. May 21, 2015

Acknowledgments

Dr. Benning, Dr. Wadford, Alice Chen, Chao-Yang Pan, and Dr. Lowney

I. Abstract

Previous studies and empirical evidence suggest norovirus outbreaks in California exhibit correlation with environmental variables and exhibit spatial spread patterns. Few studies have been done looking at what causes norovirus seasonality in temperate climates and more research is needed on the regional level. This study aims to find what relationships exist with outbreak occurrence and environmental variables in California, as well as any spatial patterns of spread or clustering of outbreaks. Spatial analysis tools were used to find any relationships between California norovirus outbreak data and environmental variables. The results showed a south to north spread of outbreaks in California and potential correlation with outbreaks and lower temperatures and higher relative humidity. More research is needed to substantiate the correlation with outbreaks and environmental variables.

II. Table of Contents

III. List of Tables

IV. List of Figures

V. List of Maps

Introduction

Overview

Norovirus is believed to be the leading cause of sporadic and epidemic gastroenteritis, accounting for roughly 50% of outbreaks worldwide. (Patel et al., 2009) Only recently has the prevalence of this virus become known. This is due to the development of reverse transcription-polymerase chain reaction; the virus cannot be cultured in a cell.

In healthy individuals the virus tends to pass without medical assistance. The virus can cause increased morbidity and fatalities in elderly, young children, and populations in developing countries. (Lopman et al., 2009) Worldwide it causes 1.8 million deaths in children under 5 years old. (Patel et al., 20009) The virus has a heavy economic burden. For the 2002-2003 season it cost the English National Health Service $184 million, for outbreaks originating in hospitals alone. In the United States norovirus illnesses, attributed to foodborne illnesses alone, is estimated to cost $2 billion annually. (Lopman et al., 2012)

There has been research to find any relationships between number of norovirus occurrences and environmental factors, such as temperature and humidity. The authors of these studies discuss the need for more investigation as they believe what influences outbreaks is a complex mixture of factors. For instance, even though the virus is typically known to peak during winter months spikes are seen in early spring. Also, individual cases are seen in summer months, however they remain isolated and don't spread like cases do in the winter. For this to happen, there must be underlying environmental or host factors that cause this. Many of the articles call for more research to see if the same environmental factors correspond with increased norovirus occurrence are present in other locations, as well as temperate and tropical settings. (Lopman et al., 2012)

The virus is generally known to exhibit winter seasonality in temperate climates, and move south to north during its season. (Inaida et al., 2013) It is not known what causes this seasonality, and what may influence the spread of outbreaks.

This paper will investigate relationships between norovirus occurrences and environmental variables through a literature review and case study of norovirus data

collected for the state of California. Finding what environmental factors influence the spread of norovirus will help with prevention measures. These measures include; deciding on what public health announcements to make, who is most at risk, and the ability to predict where outbreaks may occur. One article in particular, stressed the need to see the impact of environmental variables on the spread of the virus, due to the possibility that climate change could change the spread and occurrences of outbreaks.

Background

History and epidemiology of virus

Norovirus has a long history; in 1929 Zahorsky recognized an illness that peaked in colder months and caused vomiting and diarrhea. He first described it as the "winter vomiting disease". Kapikan, using an immune electron microscopic examination of samples, then identified the virus in 1972. It was found to belong to the *calciviridae* family. The development of reverse transcription-polymerase chain reaction, genomic sequencing, and the molecular cloning of the viruses genome in 1990 have led to a better understanding of its epidemiology. (Patel et al., 2009)

Norovirus is comprised of a non-enveloped, single-stranded RNA genome. They are very diverse and there are approximately 40 genotypes that are divided into five genogroups (GI, GII, GIII, GIV, GV). (Rohayem, 2009) GI and GII are mainly responsible for human cases of norovirus. Within the GII genogroup there are at least 19 genotypes and one of them, GII.4, is attributed to more than 85% of outbreaks. There are many variants of GII.4. (Patel et al., 2009)

The last pandemic was in 2006. It was attributed to the variant GII.4 Minerva, which has now subsided. This variant has been replaced by GII.4 New Orleans, showing that strains can be displaced. The way that these pandemics occur and subside suggests population immunity to a variant; a new variant emerges during a subsequent pandemic. Statistical models of population immunity and emergence of new GII.4 variants found increases in virus occurrence associated with low population immunity, and with emergence of new variants. (Lopman et al., 2009) Not all variants become pandemic and this is still not fully understood. This may be due to histo-blood group antigen binding patterns. (Vega et al., 2011)

Norovirus affects people of all ages causing nausea, vomiting, abdominal cramps, muscle pains, and non-bloody diarrhea. Symptoms typically subside after two to three days. Outbreaks are common in institutional settings such as long-term care facilities and childcare centers. The only treatment for patients is to keep them hydrated with solutions that contain electrolytes, and, when they are able to tolerate it, to offer food high in calories. Antibiotics and antimotility agents have not been shown to help. (Patel et al., 2009)

There has been interest in developing a vaccine, but more research is needed on the virus and immune response. (Patel et al., 2009) The other complication with vaccine development is how fast new GII.4 strains can emerge. A vaccine would need to be reformulated each time a new GII.4 strain emerged. Researchers do believe that with more investigation of GII.4 blockade epitopes that it would be feasible to quickly reformulate vaccines to the current epidemic strain. (Debunk et al., 2013) More research would also need to be conducted on the economics of creating such a vaccine.

Norovirus is considered to be the leading cause of acute gastroenteritis, and is known to have a greater effect on children and the elderly. Outbreaks typically occur in settings where people are in close quarters. Studies show that settings such as long-term care facilities are prone to outbreaks.

A meta-analysis of gastroenteritis cases was conducted to find the prevalence of norovirus. As well as looking at overall prevalence, the authors looked at prevalence across different variables such as age, setting, and developing versus developed countries. Looking at prevalence of norovirus across different variables can give insight into the ecology of the virus, and who is most at risk. (Ahmed et al., 2014)

The authors found that 18% of acute gastroenteritis cases were attributed to the norovirus. Of those cases there was little variation between prevalence across ages. Age groups were limited to less than five years, over five years, and mixed ages. The prevalence of the virus was higher in community (24%) and outpatient settings (20%) versus inpatient settings (17%). It was higher in low-morality developing (19%) and developed countries (20%) versus high-mortality developing countries (14%). (Ahmed et al., 2014)

Prevalence is difficult to quantify. Symptoms of the illness are diarrhea and vomiting for a short period of time; many people do not seek professional medical attention. Also, culturing a sample cannot identify norovirus. This means some cases are never confirmed by laboratory analysis. (*Norovirus Diagnostic Methods, 2014)*

Norovirus is most commonly under-reported in young adults, particular young men. A study conducted in Germany looked to ascertain the magnitude of under-reporting for the norovirus. The authors looked at average number of cases reported for norovirus for previous years versus the number of cases of norovirus reported during an e-coli outbreak. The thought was due to the heightened public awareness of diarrhea as a symptom of possible e-coli infection more patients with this symptom would seek professional medical attention rather than try home remedies. The authors believe that still not everyone with gastroenteritis was seen in this time period or tested for norovirus, so their under-reporting factors would be the minimum. The under-reporting factors were found to be different across age brackets, 20-29 year olds had the highest under-reporting factor (factor of 2-3) and there was minimal to no under-reporting factor for children under 10 years of age as well as adults 70 and older, and genders, males in the 20-29 year old bracket had the highest under-reporting factor. (Bernard et al., 2014)

Known transmission routes

There are many ways to spread the virus, which makes for complex chains of transmission in outbreaks. The primary mode of transmission is person-to-person and foodborne transmission. This is because the virus is environmentally stable and highly infectious. (Lopman et al., 2012) The virus is very persistent in the environment and can remain active in freezing temperatures and up to 140°F. (Hall et al, 2014)

Researchers completed a study that looked at CDC data of foodborne norovirus outbreaks from 2009-2012. Their results showed that of foodborne outbreaks 90% occurred in food preparation settings. Restaurants were the most common setting at 64%, and catering or banquet halls had 17% of the outbreaks. For outbreaks that reported factors for food contamination 70% implicated infected food workers, of these cases 54% reported bare-hand contact with ready to eat food. The authors also refer to a previous study that also points to infected food workers as the primary source of contamination.

This is an issue because food workers "…have the potential to significantly amplify community transmission of noroviruses through widespread exposure." (Hall et al., 2014)

Recommendations to reduce spread of foodborne outbreaks include following proper hand washing guidelines, avoiding bare-hand contact with gloves and utensils, following policies that prevent ill workers from working until 48 hours or more after symptom resolution, and supervision by a certified kitchen manager. These recommendations stem from observational studies that show proper hand washing is only done for 27% of activities it is recommended for, and only 16% when gloves were used. Also, one in five workers reported having worked while ill; this is due to fear of job loss or leaving coworkers short staffed. (Hall et al., 2014)

Other environmental routes of transmission include fecal-oral, vomit-oral, and even a small portion of reported waterborne outbreaks. The virus can remain infectious in water for two months. Scientists have found intact virus capsids, the protein shell of a virus, in water for over three years. It can be transferred between hands and surfaces causing a chain of transmission. (Lopman et al., 2012)

In Denmark an outbreak of norovirus was attributed to a contamination of tap water. The drinking water line was found to be broken and a nearby sewage line leaked into it. Fecal samples from persons infected and tap water were sampled. Laboratory analysis of theses samples identified the same strain of norvovirus in both samples, confirming that people had been infected from tap water. (Van Alphen, 2014)

The environmental persistence of norovirus has been documented. In one instance a concert attendee vomited at the concert hall and five days later over 300-concert attendees developed gastroenteritis. Investigators determined that people were of higher risk of developing gastroenteritis if they had been seated closer to where the initial attendee vomited. Vomiting has been found to intensify the spread of norovirus. (Lopman et al., 2012)

Outbreaks typically occur in institutional settings, such as hospitals. Reoccurrences can happen, even after sites have been thoroughly cleaned. The SARs virus was spread through a hospital in South East Asia via infected water traps. In this case, virus-laden aerosolized droplets were able to enter spaces due to defective water

traps. Gormely (2014) designed a study to determine if the norovirus could also spread in this manner.

Samples of wastewater were collected from collection drains at a hospital to determine if the building drainage system was contaminated. Samples were tested to attempt to positively identify norovirus GII strain. Unfortunately, sampling for the building drainage system airflows was considered to be ineffective, as norovirus was undetected in all the air samples. However, researchers were able to conclude that the building drainage system was contaminated. This type of contamination can cause virus-laden droplets to rise and fall in the drainage system in response to changes in humidity and airflow, and to emerge elsewhere and possibly infect a new host. (Gormley et al., 2014)

Norovirus exhibits winter seasonality similar to airborne viruses, such as influenza and measles. (Rohayem, 2009) This epidemiological feature alongside with how fast norovirus can spread in a community, leads researchers to believe that it may be airborne and spread through respiratory droplets. If this is the case it is another important route of transmission to consider when creating prevention measures. (Mounts et al., 2000)

The virus has a short incubation period, and the host remains infectious for a long time, making it difficult to track how the virus spreads. The virus can shed in high loads in stool for two weeks after infection. Only small doses are required to infect the host. Figure 1 Shows direct and indirect transmission potential of norovirus over time. (Lopman et al, 2012) helps illustrate the potential of direct and indirect transmission over time. Direct transmission is the spread of infection person-to-person. This transmission is highly infectious in the first two days then chance of infection dramatically decreases. Environmental transmission, spread of infection through contaminated food etc., can be highly infectious in the first day, but less so than direct transmission. However, environmental transmission can infect new hosts longer than direct transmission. (Lopman et al., 2012)

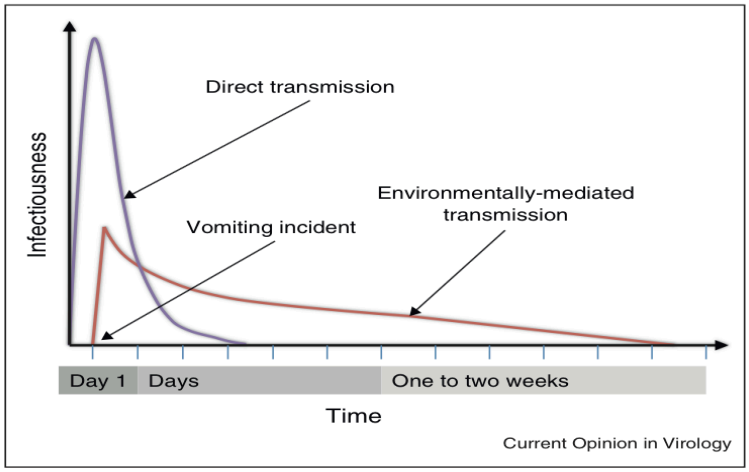


Figure 1 Shows direct and indirect transmission potential of norovirus over time. (Lopman et al, 2012)

Prevention

Due to the high environmental stability of the virus the best measures to take to reduce the spread of the virus involve cleaning contaminated surfaces. To reduce the risk of transmission the CDC recommends using a bleach solution with a concentration of 1,000-5,000ppm to clean non-porous surfaces. This solution should be used to immediately clean areas where someone has vomited or had diarrhea. Any laundry items that may have been contaminated with fecal matter or vomitus should be machine washed and dried. Avoid shaking out any contaminated laundry items, as it could cause spread of the virus. It is recommended that when cleaning any items or surfaces that may have been contaminated to wear gloves and to wash hands when finished. (*Prevent the Spread of Norovirus,* 2014*)*

Routinely washing hands with soap and water for 20 seconds can help stop transferring the virus to surfaces. Studies of the effectiveness of alcohol-based hand sanitizers are inconclusive. A study of long-term care facilities shows that facilities that use alcohol-based hand sanitizers actually have higher outbreak rates than facilities that don't use them. It is thought that the use of hand sanitizers reduces the amount of times one washes their hands, as the CDC mentions routine hand washing is important to stop transmission. (Lopman et al., 2012)

Norovirus outbreaks commonly occur in settings that have a large group of people in a common area with close living quarters. Such as hospitals, nursing homes, and cruise ships. Outbreaks in hospitals can be especially worrisome as the virus can affect patients

who are already ill, which can cause further complications to their recovery. Some measures that can be taken to stop or prevent an outbreak in a hospital is to isolate and group patients with similar gastroenteritis symptoms together, and ensure that strict sanitation measures are taking place for those interacting with them. More extreme measures that can be taken is to turn away new patients, as well as furloughing any staff who is ill for 72 hours after their symptoms subside. (Johnston et al., 2007)

On cruise ships 90% of outbreaks with diarrhea as a symptom were attributed to norovirus. Since the virus is environmentally persistent it is able to cause consecutive outbreaks, even after the vessel was cleaned. In 1975 the CDC established a Vessel Sanitation Program (VSP) with the cruise ship industry in order to combat norovirus outbreaks. The VSP sets standards for environmental sanitation and food handling. They also routinely inspect the vessels. Even with the VSP norovirus outbreaks still frequently occur on board cruise ships. An investigation of a confirmed norovirus outbreak on a cruise ship in 2009 revealed several infractions on proper sanitation procedures. This highlights how important sanitation is to prevent environmental transmission of the virus. (Wikswo et al., 2009)

Food safety is also important to consider when preventing the spread of norovirus. When ill with norovirus it is recommended to not prepare food for others, and to then wait at least two days after symptoms have ended. When preparing food the CDC recommends washing all produce before consuming, and if eating shellfish making sure to thoroughly cook them. This is because produce can be contaminated in the field, and shellfish may be harvested from contaminated waters. (Prevent the Spread of Norovirus, 2014)

A Center for Disease Control (CDC) study from 2001 to 2008 found the most likely foods to be infected with norovirus are leafy greens (33%), fruit/nuts (16%), and mollusks (13%). Testing for norovirus in foods can be costly due to lab equipment needed. Also, real time reverse-transcription polymerase techniques cannot distinguish between live and inactive norovirus. This is a problem as food may be rejected as infected due to a positive result on testing, when it has been through proper processing to sanitize it. There is a need to standardize detection methods on food so that it is uniform.

A detection strategy requires three steps; sampling strategy, RNA purification method, and a molecular detection assay. (Stals et al., 2013)

Research on appropriate techniques to kill norovirus on food is limited due to inability to culture human norovirus in lab, propagate it in vitro, and lack of suitable laboratory animals. Scientists rely on using surrogate viruses to test effectiveness of food sanitization techniques. The murine norovirus (MNV) is believed to be the most suitable surrogate. (Sanchez et al., 2011) MNV was chosen as a surrogate for a variety of reasons; the most important being that it can readily be cultured. (Hewitt et al., 2009)

Testing new ways to kill the virus in food has become more important as consumers are consuming more minimally processed foods and they need to find non-thermal ways to treat the food. One new process that does not require heat, is high hydrostatic pressure processing. Initial research shows that it can reduce the amount of murine norovirus present in food, as well as human norovirus. The only problem is that some manufacturers add calcium to food to increase its firmness, and this can make the virus more resistant to pressure. (Sanchez et al., 2011)

Lab analysis and confirmation

Real time reverse-transcription PCR (RT-PCR) can be expensive and time consuming. Dr. Wadford and Mr Chao-Yang Pan at the California Department of Public Health detailed the time and cost associated with norovirus sampling. A single sample would cost $35 without including equipment costs or a microbiologist's time. If those factors were included it would cost on average $500, this is because it takes two days to determine norovirus genotype. However, a microbiologist can run multiple samples at a time, anywhere from 24-48 samples. If multiple samples are run it can bring the cost down to $50 on average. This will provide results that determine the virus's genotype and a phylogenetic analysis. This is important to track new strains of the virus. (Personal communication, April 14, 2015)

The equipment used for sampling is expensive, and multiple pieces are needed. There are two different machines that can be used for nucleic acid extraction; they range from four to eight thousand dollars. For these machines it can cost about eight dollars a sample. Other equipment needed include real time instruments, $8,000, and conventional thermal cyclers, $3,000. Reagent tests can cost one to two dollars, and several of these

tests need to be run, (GI, GII, MS2 or region C and D) (Dr. Wadford and Mr. Chao-Yang Pan, April 14, 2015)

Other expenses include PCR purification kits, approximately $1 per reaction and up to two done per sample. Sequencing of samples, up to 4 reactions are needed per sample costing between $3 and $5. (Dr. Wadford and Mr. Chao-Yang Pan, April 14, 2015)

The time cost associated with real time PCR is four hours, then another four hours for the conventional PCR result. Additionally it can take two hours to run a gel and get it ready to be sent out. When sequences come back, it could take half day to analyze a whole run. The results also need to undergo quality control, which can take another hour. After all this the sequences can then be read and uploaded. (Dr. Wadford and Mr. Chao-Yang Pan, April 14, 2015)

The current method of detection for norovirus uses RT-PCR techniques. This technique was made available once the sequence of the norovirus was known, improving the ability to detect the presence of the virus. Due to the strain diversity multiple primers are used. Immunoassays can be used in outbreak settings. These tests are quick to perform, don't require extensive lab equipment, and are sold commercially. They are less sensitive than real-time RT-PCR but when used with multiple samples can still provide useful results to confirm if an outbreak is attributable to norovirus. (Glass et al., 2009)

Table 1 depicts the advantages and disadvantages of different tests used to diagnose norovirus in a patient. (Kirby et al., 2012) A clinical test checks to see if patient admitted has symptoms that match certain criteria. Such as the Kaplan criteria "a mean (or median) illness duration of 12 to 60 hours, a mean (or median) incubation period of 24 to 48 hours, more than 50% of people with vomiting, and no bacterial agent found." (*Responding to Norovirus Outbreaks,* 2013) Electron microscopy was how the virus was first identified. It is a quick test to run, but lacks the sensitivity that polymerase chain reaction has. There are two methods of immunological tests, or immunoassays, available. They are ELISA and immunochromatographic testing. As discussed above they are rapid tests that do not require lab equipment, but they lack sensitivity and specificity of other tests, and are useful in an outbreak. Real time reverse-transcription polymerase chain reaction continues to be the gold standard for testing, but requires additional laboratory

equipment and training. Luminex is a type of multiplex polymerase chain reaction that also allows for testing of other gastrointestinal pathogens in the same test. (Kirby et al., 2012)

Table 1 Advantages and disadvantages for various methods of testing for norovirus (Kirby et al., 2012)

Test		Advantages	Disadvantages
Clinical test	Sporadic cases	Rapid Point of care No consumables	Clinical training is required Inconsistent results
	Outbreaks	Useful in the absence of a laboratory	Moderate analytical sensitivity Data collection required
Electron microscopy		Rapid Identifies a range of viral pathogens	Training is required Electron microscope required
Immunological	ELISA	Specific Specialist equipment not required	Moderate overall analytical sensitivity Poor sensitivity for GI noroviruses
	ICG	Rapid Specific Specialist equipment not required Single sample used Potential for point of care	Moderate overall analytical sensitivity Poor sensitivity for GI noroviruses
RT-PCR		High analytical sensitivity and specificity Can be multiplexed Quantitative	Clinical specificity reduced by high analytical sensitivity PCR equipment required
Luminex		High analytical sensitivity High analytical specificity Can be multiplexed Quantitative	Clinical specificity reduced by high analytical sensitivity Luminex equipment required

To confirm the presence of norovirus in patients the CDC recommends collecting a stool sample. This collection method can lead to difficulties in collection and storage. Rectal swabs are another method for collection; their diagnostic performance is still being investigated. The ease of use for rectal swabs leads researchers to believe that it would increase the number of specimens collected, providing more available data. (Arvelo et al., 2013)

Recent studies have been looking at the performance of rectal swabs versus stool samples for lab analysis. Rectal swabs can be advantageous for sampling patients that are too dehydrated to produce a stool sample, or for recently deceased patients. To compare

the two sampling methods researchers collected stool samples and rectal swabs from patients displaying symptoms resembling norovirus. The samples were then tested for norovirus using real-time reverse transcription PCR. (Arvelo et al., 2013)

The results for stool samples and rectal swabs differed, but both had the same diagnostic performance. Eight patients tested positive by rectal swabs only, and six patients tested positive by stool sample only. For patients that tested positive for norovirus only 36% had concordant results with the two specimen types. Simultaneously to testing for norovirus the researchers also tested for rotavirus. In comparison to norovirus patients positive for rotavirus had 84% concordant results for the specimen types. This shows that neither method of sampling for norovirus should be considered optimal. In an outbreak setting, where stool sample collection and storage may be difficult, rectal swabs could still be used to determine if it is attributable to norovirus, due to multiple patient samples. (Arvelo et al., 2013)

Since norovirus is environmentally stable the virus can be detected in the environment. Waterborne outbreaks of norovirus can be determined through sampling of water. Detection of norovirus in water can be difficult, as it requires large quantities of water, due to waters low viral content. (Van Alphen et al., 2014) The viral content is low due to dilution and because without a host cell the virus cannot replicate. (Verheyen et al., 2009) Even though the viral content in water may be low it takes less than 10 viral particles to infect a new host. (Patel et al., 2009)

There are defined methods for extracting norovirus from shellfish for testing. Other food items may be collected for sampling as well. Swabs of environmental samples can be analyzed as well but the results are variable, and should be interpreted with caution. (Specimen Collection, 2013) Swabs of surfaces can be useful when investigating outbreak reoccurrences to see if surfaces have been cleaned properly.

Surveillance and Management

Surveillance and data sharing networks

The studies all draw the same conclusion: a complex mix of environmental and host factors play an important part in how the virus spreads. To learn more about how the virus works, better data gathering and sharing is recommended. Data gathering is

accomplished through different surveillance methods There are two different types of surveillance techniques used to gather and analyze data on outbreaks. They are syndromic surveillance and lab surveillance. Once data has been gathered there are a variety of different ways to share, store, and analyze the data electronically. Figure 2 depicts a model of how surveillance data should be analyzed for early outbreak detection. It outlines the steps that should be taken after statistical analysis of data collected signals something unusual. From there epidemiologists can look through the data and statistics to determine the probability of an outbreak or if there was an error in data reporting or processing. If the probability of an outbreak is high further investigation is needed to look at the cause of the increase in cases. (Buehler et al., 2004)

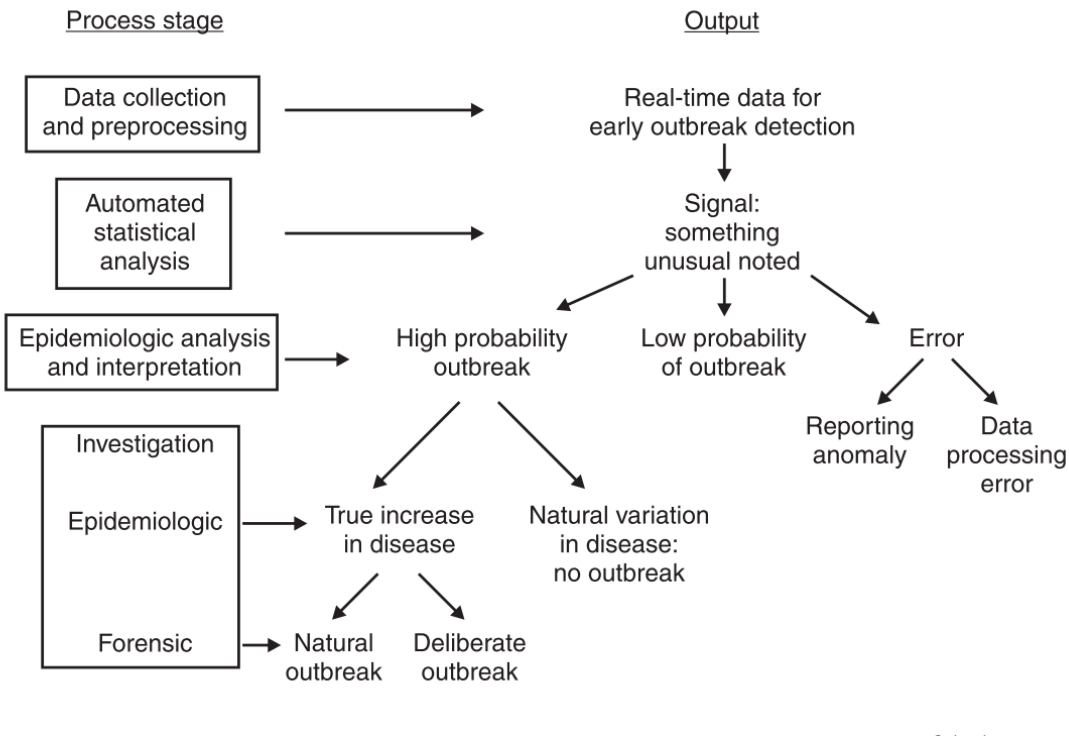


Figure 2 Process model for early detection (Buehler et al., 2004)

It is important to investigate the effectiveness of the surveillance system for outbreak detection as the development and management of these systems can be costly. The CDC has developed a four-part framework to help determine the surveillance systems usefulness. This framework should also help offer insight on how to create a surveillance system or improve upon an existing one. (Buehler et al., 2004)

Syndromic surveillance

Syndromic surveillance covers a variety of survey methods that help to provide an early detection for outbreaks. Figure 3 illustrates how it can detect outbreaks days before lab confirmation. It can help researchers learn more about outbreak patterns, magnitudes, and trends. The CDC defines syndromic surveillance as "an investigational approach where health department staff, assisted by automated data acquisition and generation of statistical alerts, monitor disease indicators in real-time or near real-time to detect outbreaks of disease earlier than would otherwise be possible with traditional public health methods." (Henning et al., 2004) These disease indicators can be anything from absentee logs to over the counter drug sales.

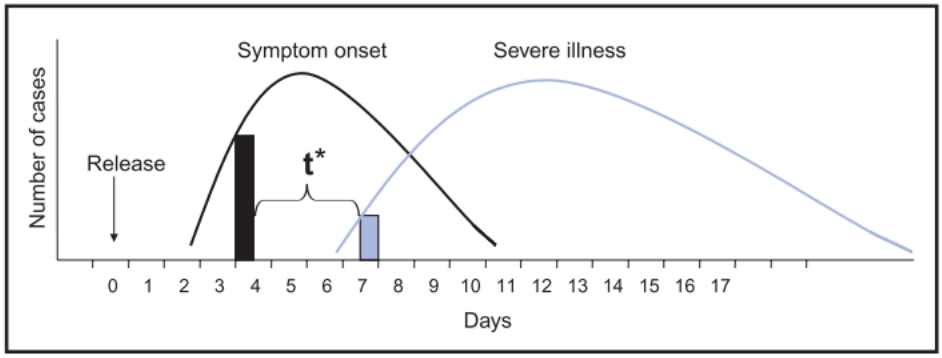


Figure 3 This graph shows how syndromic surveillance can help with early detection of an outbreak. (Henning et al., 2004)

One example of syndromic surveillance used to predict outbreaks of norovirus is analyzing word patterns in search engines. Websök is a system that analyzes data, created by search queries, from the Stockholm online health portal. This system was initially created to monitor influenza-like illnesses. (Edelstein et al., 2014)

A study was conducted to see if this same system could be utilized to track norovirus outbreaks. The authors tailored the system to track the terms "vomiting" and "winter vomiting disease". They found that peaks in searches containing those words came before laboratories reported norovirus outbreaks. Using the term "winter vomiting disease" showed a higher correlation with occurrence of norovirus, as "vomiting" was too broad. This system only detects overall trends and season onset. It cannot be used to look

at severity of outbreak, as one search for the keywords does not equate to one case of norovirus. (Edelstein et al., 2014)

The authors concluded that the use of the Websök system helped to earlier detect the onset of the norovirus season. The system cannot replace laboratory data, but if used in conjunction it is helpful as an early alert system for health care professionals. This can be useful to help prepare infection control measures. The system is also low cost to implement, and only requires a local health related search engine. A local search engine should be used since norovirus is thought to have a correlation with climate. (Edelstein et al., 2014)

Surveillance van also be used on a global scale to find emerging strains of norovirus. This information can be useful as new strains can become pandemic, early warning can allow public health officials to take appropriate measures to treat and prevent outbreaks. A new variant was suspected as the United Kingdom, Japan, and Netherlands reported seeing more cases than in previous seasons. Data on norovirus cases was uploaded into an international molecular surveillance database, called NoroNet. From there it became clear that the increase in norovirus cases was due to a new variant, of a genotype II.4 norovirus, first seen in Australia, the variant was named GII.4 Sydney. The emergence of a new epidemic variant is seen every two to three years. (van Beek et al., 2013)

Lab surveillance

Lab surveillance looks at lab results uploaded into electronic reporting systems. These systems have the ability to mange and analyze data efficiently. This data can alert health care professionals to potential outbreaks if increased number of occurrences of a virus are reported. The use of electronic reporting systems has been shown to increase the amount of data entered and cases reported, leading to better analysis of the disease. (Samoff et al., 2013)

In the United States the CDC launched an electronic platform to collect data on norovirus called CaliciNet. The goal of CaliciNet is to help with prevention measures and to better analyze norovirus. Data is collected from participating public health laboratories on the federal, state, and local level. In 2014 28 states, 33 laboratories, and the District of Columbia have received the certification necessary to participate in CaliciNet. Figure 4

shows a map of states participating in CaliciNet. If a laboratory is not certified they can send their samples to a CaliciNet Outbreak Support Center for norovirus typing. Data collected is analyzed to help identify outbreaks and find the potential source. (Reporting and Surveillance for Norovirus, 2015) To become certified a laboratory is undergoes a laboratory panel test and is evaluated on data entry and analysis of sequences. Once certified labs must pass an annual proficiency test. (Vega et al., 2011)

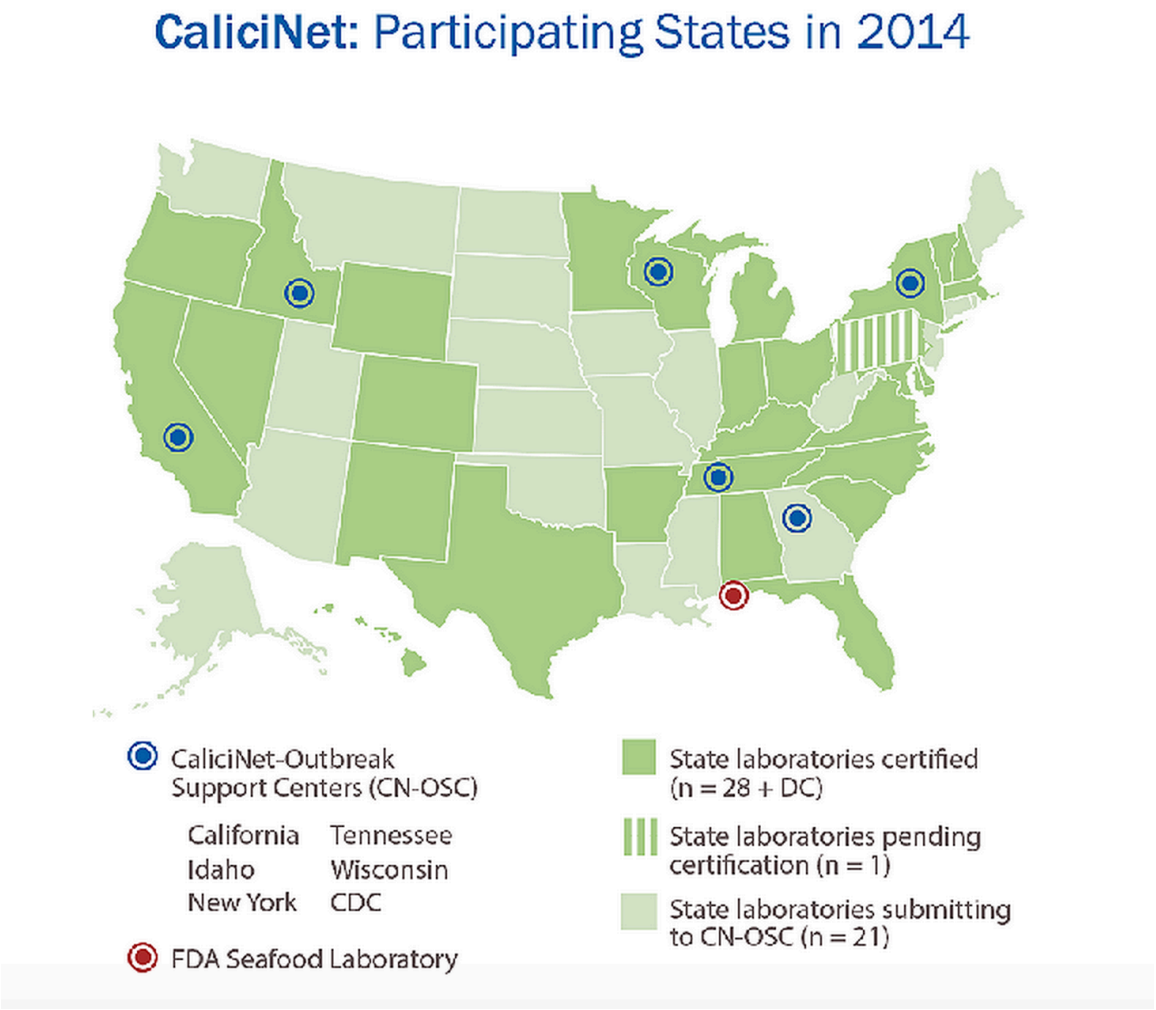


Figure 4 States participating in CaliciNet (Reporting and Surveillance for Norovirus, 2015)

This is important as if the virus was caused by consumption of a food item, such as shellfish, the public can be warned about the dangers of consuming it. The data can also help to identify any new strains of norovirus that may be emerging. New strains can become pandemic and with early warning health officials can prepare for the influx of patients as well as create public health messages to help prevent transmission.

Information from CaliciNet in the winter of 2009-2010 helped identify a new GII.4 variant, GII.4 New Orleans, which became the predominant GII.4 strain. (Vega et al., 2011)

There are many different electronic reporting systems used throughout the country and world. Creating a singular platform that could be used among all participating stakeholders could increase the ability of data sharing and analysis. One such idea is to create health information organizations (HIOs). Its structure would help create a unified interface, format, and terminology. Additional reporting of norovirus can help to find disease patterns in communities; this can help guide public health messages and help identify outbreak sources. HIOs will be able to handle specific searches, which can help researchers identify trends. Finally, automated HIOs will help labs voluntarily report outbreaks efficiently. (Shapiro et al., 2011)

Reportable Diseases

The CDC does not classify norovirus as a reportable disease. This means that singular occurrences do not need to be reported but outbreaks do. An outbreak is when the occurrence of disease is greater than expected in an area, community, or season. (Disease Outbreaks, 2015) For norovirus an outbreak is when there are two or more occurrences that have a common exposure and is suspected, or laboratory-confirmed, to be caused by norovirus. (Reporting and Surveillance for norovirus, 2015)

Diseases that are deemed reportable diseases by the CDC are considered to be of great public health importance. The reporting allows for data collection that helps analyze the disease occurrence. (Reportable Diseases, 2015)

If norovirus was classified as a reportable disease each confirmed case would be reported to the CDC. This could create an invaluable wealth of data that could lead to in depth analysis of the disease that leads to better control and preventions measures. It could also help to evaluate the effectiveness of prevention measures. (Buehler et al., 2004) Many of the studies done on norovirus outbreaks lack the data needed to create an in-depth analysis.

Modeling of Outbreak Spread

Why use models

Understanding a virus through research and surveillance can help to create a predictive model. A predictive model can simulate the potential spread of a virus. To create one research must be done to see what factors affect the virus and how. (Hyder et. al., 2013) It is also important to research the host population demographics and geography of study area. Once a model is created for a specific disease it can be used to help health officials prepare for outbreaks by predicting where the virus will spread and at what rate. (Modeling Infectious Diseases, 2014)

Typical models used when studying disease spread are agent-based, meaning that it models host behavior in a community. This is important as it takes into account the opportunities a host can have to either spread or be exposed to the disease. This is also quantified by looking at how the disease spreads, person-to-person or airborne etc. Scientist's model host behavior by making assumptions on how hosts interact in a community, this can be done by looking at demographics. For instance a community in a rural setting may have less host-to-host interaction, opportunities to spread or contract the disease, versus an urban setting. (Modeling Infectious Diseases, 2014)

Once a model has been created its inputs can be changed to look at different scenarios. It can be used to test prevention measures, such as vaccination or quarantine. Or it can model how a virus will behave in different settings, rural or urban. Another important input that can be modeled is the contagiousness of the disease; this can help to model the effect of different virus strains. It is important to note that when creating preventative measures results usually include a variety of different prevention measures. These prevention measures may also be implemented in different time intervals. (Modeling Infectious Diseases, 2014)

Scientists create multiple models for the same disease and setting to see if the results coincide. This is because no model can definitively reflect real world scenarios they are only as good as their inputs. If multiple models give similar results on how the disease may spread it gives confidence to the prediction. (Modeling Infectious Diseases, 2014) Model precision is important as models that do not accurately predict "peak week, intensity, and duration…" can have high economic consequences. If a model underestimates the duration of an epidemic it can cause vaccine and other resource shortages. (Hyder et al., 2013)

The researchers at Models of Infectious Disease Agent Study (MIDAS) are considered the leading experts for creating models for disease spread. They formed in 2004, and are funded through the National Institute of Health. Part of their success in creating predictive models is due to their understanding that it requires research in many different fields. They employ researchers in "epidemiology, infectious diseases, computational biology, statistics, social sciences, physics, computer sciences and informatics." (Modeling Infectious Diseases, 2014)

Influenza Model

Multiple predictive models have been created for influenza. These models will be key to consider when creating a model for norovirus since they share some similarities. Influenza exhibits winter seasonality in North America, and can persist throughout the year sporadically. There are several types of influenza, and within those different subtypes. Influenza A is one of the most common types to infect human hosts, and it evolves every year. Hosts can develop immunity to a flu strain. Once infected a host develops symptoms within three days, and can shed the virus for up to ten days after symptoms present. Young children and the elderly are considered to be more susceptible to infection. (Gunder and Dadig, 2010)

Influenza can cause death and can have a heavy economic and social cost during epidemics. Modeling of influenza spread throughout a season can help policy-makers reduce the negative effects of influenza by creating prevention measures. There are vaccinations available for influenza and modeling can help decide the number of vaccines needed for the season, as well as where and to who they should be distributed. If a model predicts an intense epidemic in some areas it can be used to determine social distancing measures, such as school closures and quarantines. (Hyder et al., 2013)

Models can also predict the timing of the epidemic, when it begins and ends as well as the peak week. This can be important information for health care officials for timing of vaccinations, and public awareness campaigns. All of these prevention measures can be tested in the model to find the most effective plan for reducing the impacts of the flu season. Researchers are also integrating influenza transmission

and climate models to predict how climate change will affect the burden of illness. (Hyder et. al., 2013)

Models of influenza incorporate a variety of factors that include host behavior and environmental variables. It is important to study these factors as models give additional weight to factors that are deemed to have a greater influence on influenza. Researchers are constantly calibrating models and fitting new perturbation factors, such as vaccination coverage, to help keep predictions accurate. (Hyder et al, 2013)

Host behavior is one of the most important factors to be considered when creating an influenza model, as hosts are what facilitate the spread. How hosts behave in different areas and during different times of year remains the same regardless of virus being studied. Host behavior data from influenza models can be used to create norovirus models. Some changes would have to be made as how hosts spread the virus is different and prevention measures implemented on hosts will vary by virus.

Absolute humidity is one of the environmental variables believed to have a correlation with influenza's seasonality. Research that looks at the relationship between influenza and absolute humidity show that influenza flourishes at high and low absolute humidity, but survival is limited in moderate humidity. This can explain why influenza is seen in fall and winter in temperate climates. As these times can exhibit the highest and lowest humidity. (Shaman, Goldstein, & Lipsitch, 2011)

Researchers looked at vitamin D as a factor in influenza spread, as vitamin D can contribute to host immunity and is believed "to be the underlying source of observed influenza seasonality in temperate regions." The results showed that that vitamin D was not the cause for influenzas seasonality in temperate climates. They did mention that it could possibly contribute to occurrences of influenza. (Shaman et al., 2011) Even if vitamin D is not a primary contributor it could still be an important factor in creating a more precise predictive model.

Norovirus Model

There is no current complex predictive model for norovirus outbreaks. Some models have been created that look at predictive qualities of individual variables. One study in England looked at how changes in relative humidity and temperature over a certain time period can influence outbreak occurrence. The study found that a 1°C increase in temperature over 35 days influenced norovirus by reducing occurrences of norovirus by 15% (Lopman, 2009) These predictive models while limited are useful for creating basic predictions.

Their needs to be extensive research done to find what factors influence the spread of norovirus. To find factors that may influence norovirus we can start by looking at what influences the spread of other diseases with the same seasonality. There have been some studies that look at geospatial patterns of spread and environmental variables that can contribute.

If there is a correlation with environmental variables model will be useful to see how climate change can affect norovirus spread. Since little is known about the epidemiology of norovirus making conjectures about how climate change may impact norovirus outbreaks is difficult. It is believed that climate change will affect viral infections in multiple ways. It can change how the virus is transmitted, host ecology, and cause socio-economical changes that can affect the host population. (Rohayem et al., 2009) It is important to study how predicted changes may affect norovirus outbreaks.

The winter seasonality of norovirus, and its correlation with lower temperatures, may be affected by climate change. To study if this will be the case increased multi year analysis of norovirus occurrences in one area should be performed. This analysis should also take into account specific locale environmental variables. The results could show if there is a shift in norovirus seasonality over the years, in reference to environmental variables. These results could be extrapolated to see how climate change will affect norovirus seasonality (Rohayem et al., 2009)

Extreme weather events caused by climate change could potentially increase the number of outbreaks. Flood events can create an outbreak due to the high possibility of water contamination. Other natural disasters can cause the need for refugee camps, which

creates large groups of people in a small area, creating the opportunity for fast and widespread transmission of the norovirus. (Rohayem et al., 2009)

Geospatial patterns

Studying the spread of outbreaks to find spatial patterns can help create effective monitoring and surveillance plans, as well as prevention measures. If the norovirus has spatial diffusion patterns it might be possible to predict where outbreaks may occur. This could help to give advanced notice to local municipalities and give them time to prepare prevention measures that may stop a potential outbreak.

Not many studies have been done on spatial analysis of norovirus occurrences or outbreaks. This may be due to lack of available data. Increased surveillance of norovirus and sharing of data relating to norovirus could help to facilitate more spatial analysis. In discussion with Dr. Wadford, Chief of the Respiratory and Gastroenteric Diseases Section at the California Department of Public Health, she mentions that in California it seems that outbreaks appear in the south then work their way north. However, there has not been an analysis to prove this theory. (Personal communication, February 10, 2015)

One of the spatial analyses of norovirus showed a south to north pattern. A study of the viruses spread through Japan, through multiple seasons, showed a south to north migration indicating that the spread of the virus may be related to climate. The southern region is more temperate whereas the north is much colder. The virus did not peak in the coldest month in Japan leading the authors to believe that there are more climatic factors involved aside from temperature. They did note that each area of Japan has different levels of humidity; Northern Japan has less humidity than southern areas. (Inaida et al., 2013)

Another spatial analysis of norovirus in the United Kingdom found no spatial patterns for norovirus. This study used telehealth data, for vomiting symptoms age five and over, to do their analysis. The authors noted that the lack of spatial correlation for norovirus might be due to the inconsistencies in availability of regional level outbreak data. Also, not including the five and under age group may have resulted in fewer data points as young children are susceptible to norovirus infections. This age group was removed as rotavirus is very common in this age group and has similar symptoms. (Cooper et al., 2008)

More studies need to be done to ascertain whether or not norovirus outbreaks follow a spatial pattern or exhibit clustering in certain areas.

Environmental variables

Until recently the correlation of norovirus occurrence and environmental variables has been anecdotal. The advancement in detection techniques has led to better confirmation and reporting of the virus, which in turn has increased analysis of the virus.

Most of the current research looks at its correlation with temperature, as the virus commonly occurs during the winter months. This has been confirmed by research looking at multiple seasons of outbreaks, in eight countries, showing that the low point for disease reports was in the warmer months. Outbreak peaks didn't always occur in the same month every year. (Mounts et al, 2000)

Another study reviewed norovirus data uploaded to CaliciNet, a database of norovirus occurrences in the US. The analysis showed a peak in January and more occurrences in winter and early spring. (Vega et al., 2011)

It is unlikely that only temperature plays a role in increased occurrences. In England and Wales multi-year norovirus data was looked at to find its correlation with temperature and humidity. The results showed that there were more occurrences of norovirus during cold temperatures and lower humidity. Further studies that look at rainfall and UV were recommended, as UV may help account for host behavior such as more time spent indoors in winter. However, they recognize analysis may be difficult as those factors can be highly localized and may be difficult to correlate to available national data. Rainfall was not believed to be associated with incidences of norovirus in their initial analysis, but they were not able to look at local rainfall patterns or extreme events. (Lopman et al., 2009)

The overall research does conclude a positive relationship between norovirus occurrences and lower temperatures and humidity. Most of the studies all stress the lack of regional data available for norovirus occurrences and environmental variables. The incorporation of environmental variables in the analysis is important to see what role they play in how outbreaks spread.

California Case Study

Introduction

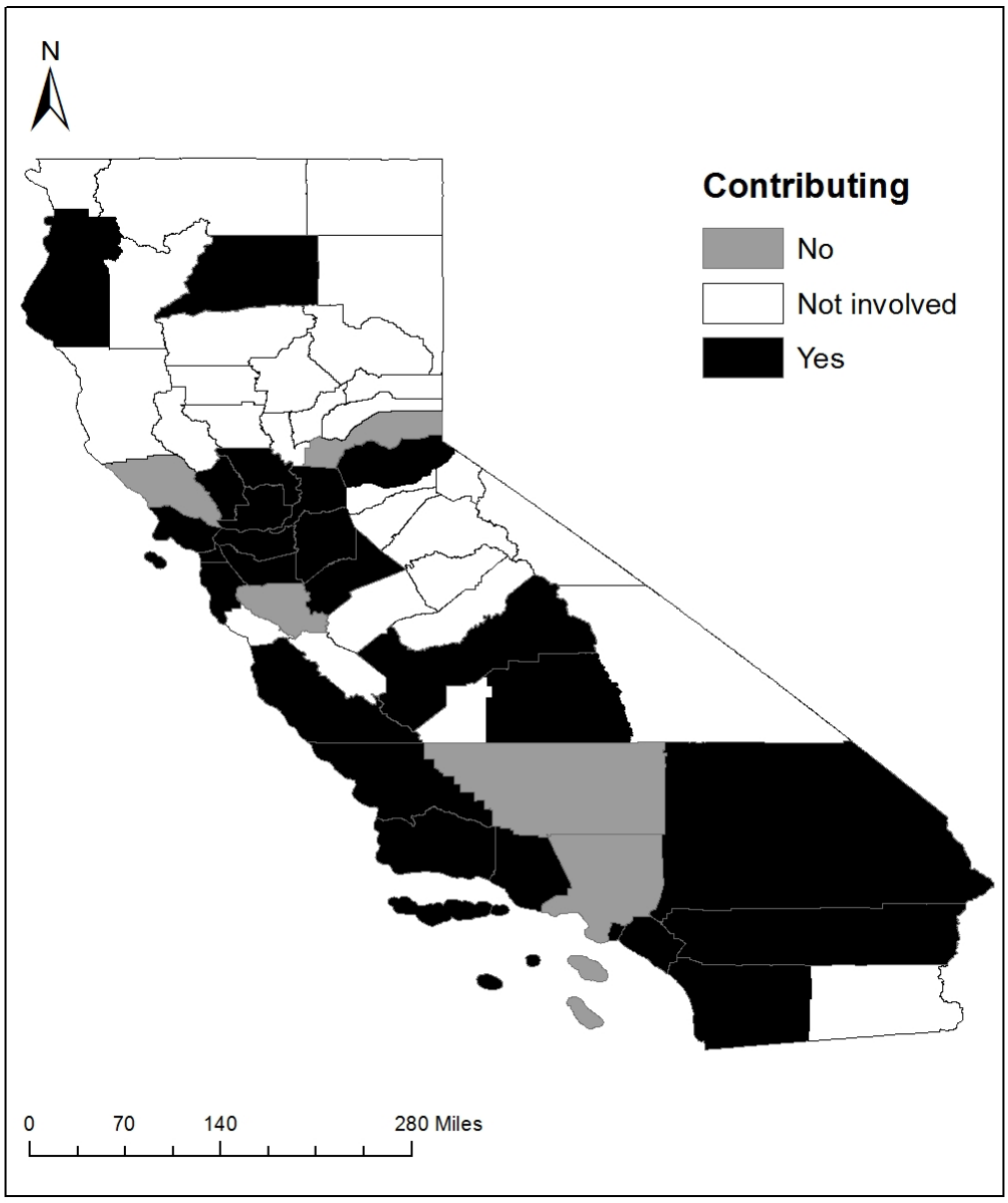
This study hopes to find spatial patterns and relationships with environmental variables for norovirus outbreaks in California. The outbreak data to be analyzed was collected by the California Department of Public Health. This dataset contains information on where the outbreak occurred, what the location type was, date of outbreak, as well as genotype.

The data will be analyzed using R Studio and ArcGIS. The results from this study can be used to learn more about the epidemiology of the virus and help create prevention measures. If spatial patterns and correlations with environmental variables are found it could be used to create a predictive model for norovirus outbreaks in California.

Challenges

There were many challenges in acquiring and analyzing the data. All counties that had available data on norovirus outbreaks were contacted and asked if they would be willing to release their de-identified data on norovirus outbreaks. Some counties were not willing to release their data. These counties include; Santa Clara, Sonoma, Kern, Los Angeles, and Placer. Long Beach is considered separate from Los Angeles County and they agreed to release their outbreak data.

In addition to counties not releasing data some are not able to participate in norovirus data collection. Map 1 depicts which counties participated, which declined, and which ones were not able to due to inability to perform RT-PCR. This is because the cost of the equipment, time, and proper training of lab analysts needed to perform RT-PCR to test for norovirus can be cost prohibitive. Twenty-four counties participated in the study, twenty-nine counties could not participate, and five counties declined to participate.



Map 1 California counties contributing to study

Many of the counties that are not able to participate have some of the lowest populations in California. Even though these counties may have a low population density they still make up a large portion of California's population, partially due to the fact that over half of California's counties could not participate. This lack of data makes spatial analysis of norovirus difficult, and may mean that key patterns of how the virus moves may be missed. Figure 5 shows the percentage of California land area that is represented in this study. Slightly less than half of California is represented in this study, 46.29%. Even though less than half of the total area of California is included in the study over 50% of California's population was included, see Figure 6. (California Counties by Population, 2015)

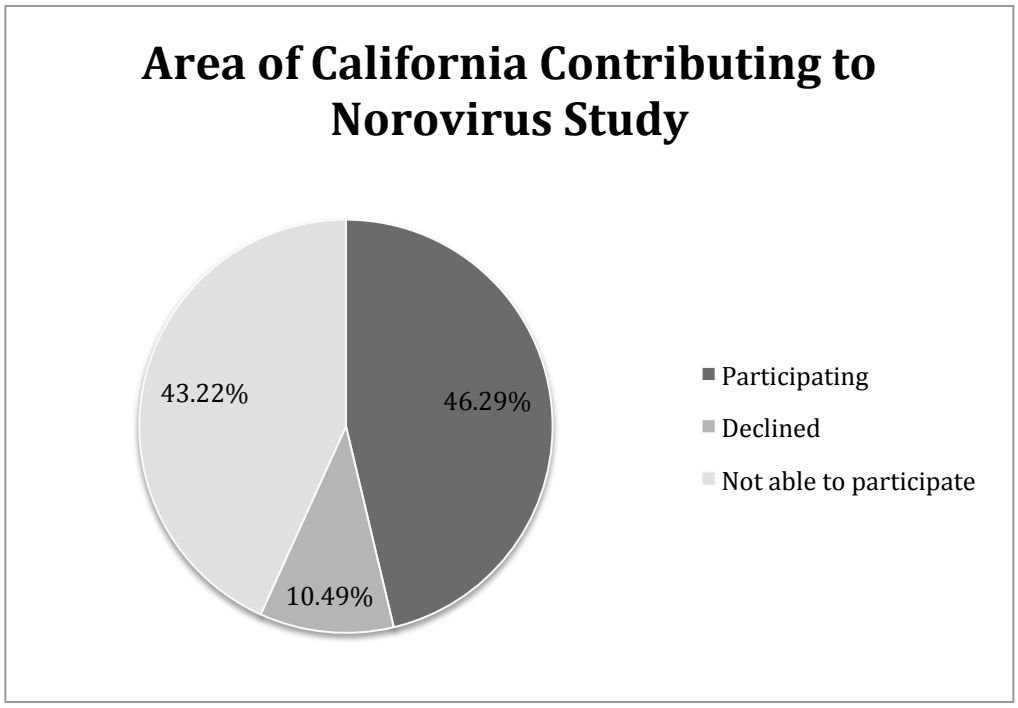


Figure 5 Area of California contributing to norovirus study

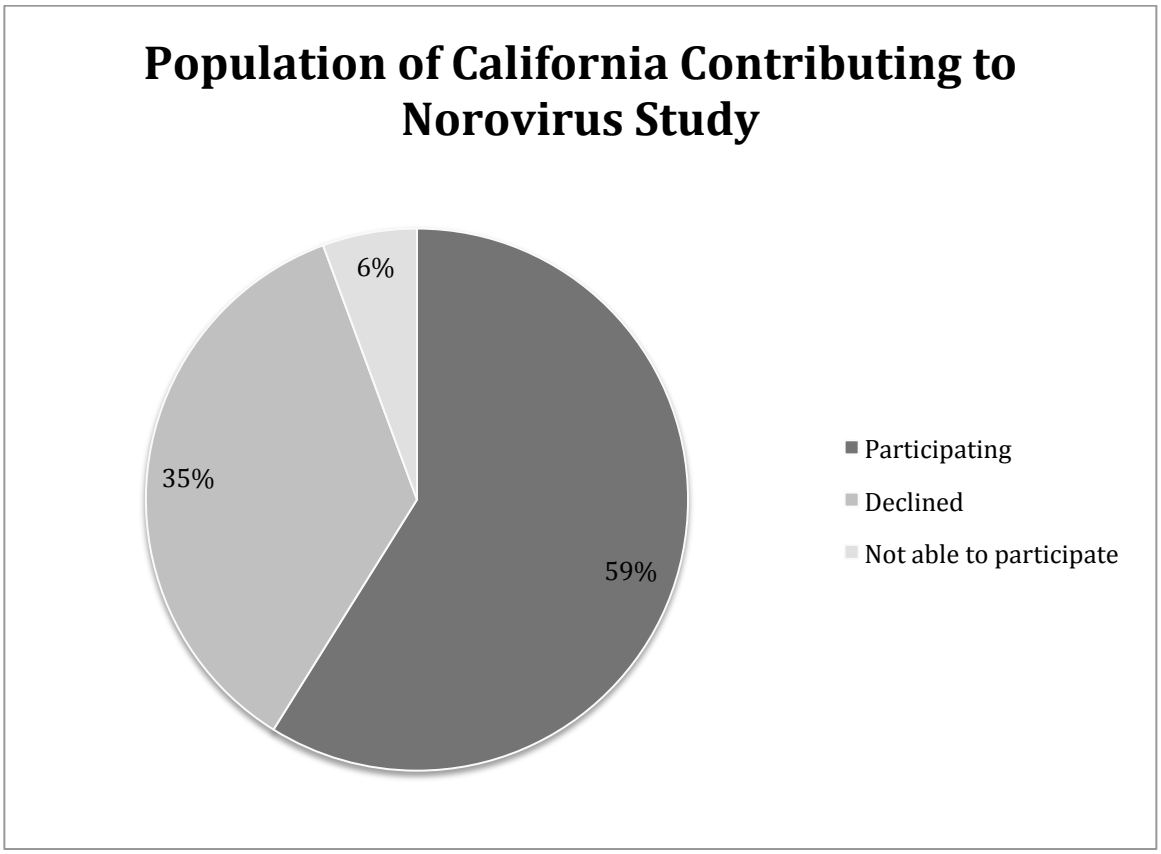


Figure 6 Population of California participating to norovirus study

The data provided from participating counties has some fields that were not populated. Many of the individual cases lacked values for genotype testing in specific regions.

Since norovirus is not classified as a reportable disease it should be assumed that some occurrences are missing from the data. If only one case is reported to health care officials it does not count as an outbreak occurrence, and may not be reported and included in datasets. Many cases also go undocumented, due to its symptoms many people choose to not seek medical care. As seen in the study on the under reporting of norovirus illnesses in Germany under reporting factors can be as high as two to three. (Bernard et al., 2014)

Gathering historical environmental variable data presented challenges. For the data used in this study a weather station per county was used. The weather station was chosen simply by whichever station had available historical data. However, many areas in California have unique microclimates and this may not be an adequate representation of the weather conditions during the outbreak. For instance, San Francisco weather can

change dramatically from neighborhood to neighborhood. San Francisco County's closest weather station, KSFO, is a considerable distance away from the heart of downtown and has a dramatically different climate. This may obscure correlations between outbreak occurrence and environmental variables.

Methods

To verify that the outbreak is attributed to the norovirus the California Department of Public Health and affiliates collected stool samples and used a real-time reverse transcription-polymerase chain reaction (RT-qPCR) assay. Different primers were then used to determine the genotype of norovirus; the virus was tested at 3 regions.

The dataset contains columns with headings for; outbreak number, county, outbreak date, ID, setting, other information, region D genotype, region C genotype, region B genotype, and transmission. The outbreak number column contains an identification number unique to the outbreak, whereas the ID number contains an identification number to the individual case. The setting column has information, when known, on the type of setting the outbreak occurred in, most commonly long term care facilities. Other information has more specific information where the outbreak occurred. The transmission column has information on how the virus spread, such as food borne or person-to-person. The setting and other information columns were sparsely populated.

The dataset was cleaned to check for errors, and to see what data may be missing. All questions on missing or incorrect data were sent to the California Department of Public Health for review. The data was condensed for analysis by removing each individual case and compiling that data into a single row detailing the outbreak. Using case data a new column was created showing the lab confirmed cases of norovirus. Taking the sum of individual cases recorded for an outbreak created this column. All data will be given coordinates to their respecting counties centroid. This is because some counties did not give data for the specific location of the outbreak to protect privacy. Using county centroids instead of exact outbreak location may obscure spatial patterns and correlations with environmental variables.

Historical weather data for outbreak locations, containing temperature, relative humidity, wind speed, and rain, was sourced from several sites including, Weather Underground or U.C. Agriculture and Natural Resources. This data was then amended to

an attribute table. The data amended with weather underground was found using the weatherData package in R. The weather station used for each location will be recorded in the metadata. Data used from U.C. Agriculture and Natural Resources was copied in its comma separated value form and inputted. The weather variables used in this study were chosen based on literature review as well as what was actually available. The column headings for this table are; county, weather station, date, precipitation inches, temperature maximum (F), temperature minimum (F), temperature average (F), wind speed (mph), relative humidity maximum (%), relative humidity minimum (%), and relative humidity average (%). The weather dataset was amended to the outbreak data to create one attribute table that can be used with ArcGIS tools.

The outbreak point data will be displayed on a map of California. The map was built using shapefiles of California county lines. This shapefile was sourced from the US census bureau and contains county lines from 2010. A shapefile of sub counties was used to find the area of Long Beach. Other data sourced from the US census bureau was population by county in 2010. This data was joined to the county shapefiles and can be used to look at population density.

To look at how the virus spreads spatially in a season spatial analysis tools available with ArcGIS will be utilized. A technique used to analyze the outbreaks in Japan was a geostatistical method called kriging to create a static map that shows the spatial distribution of norovirus cases over week intervals. Kriging is an interpolation method that models "…the spherical spread by geostatistical estimation of the point based data." (Inaida et al., 2013). Other tools that will be used include hot spot analysis to find clustering, as well as ordinary least squares to look at relationships between outbreaks and environmental variables.

The geostatistical analysis function will be used with the combined outbreak and weather dataset to find basic statistics on variables. Charts will be created to display relevant information for the state as well as on the county level.

Results

Environmental variables

Temperature

The temperature statistics were created giving equal weight to each outbreak. There were 695 values used. This data was used to create a q-q plot, Figure 7, to look at its distribution; average temperature was plotted against count of outbreaks. The q-q plot demonstrates the data is normally distributed.

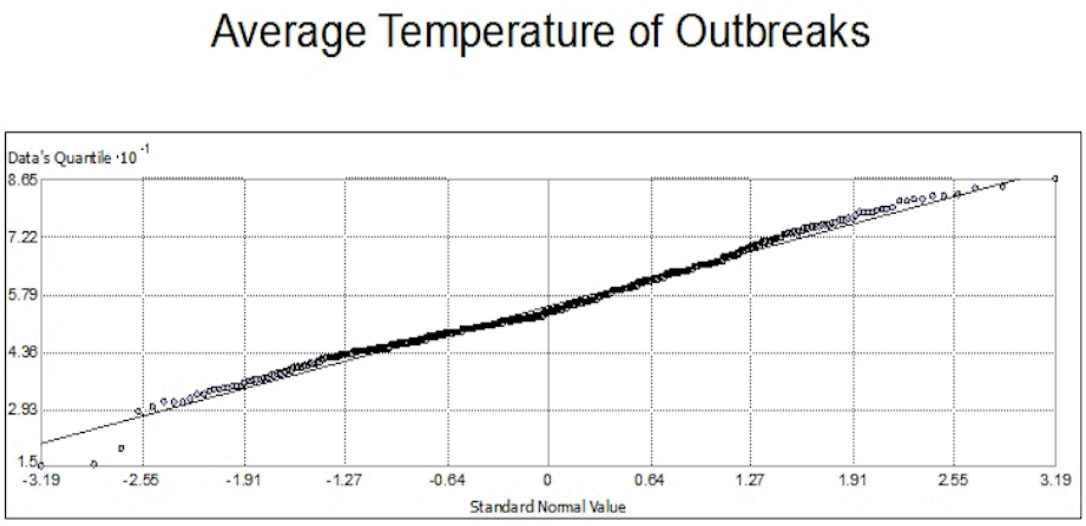


Figure 7 QQ plot of average temperature of outbreaks

Since the data follows a normal distribution further analysis is needed to find what average temperature outbreaks typically occur during. Figure 8 is a histogram of average temperatures of outbreaks, using a binwidth of 5°F. There appears to be a right skewed normal distribution of average temperatures. The mean average temperature is 54.61°F ± 0.8°F, with a minimum of 15°F and a maximum of 86.5°F.

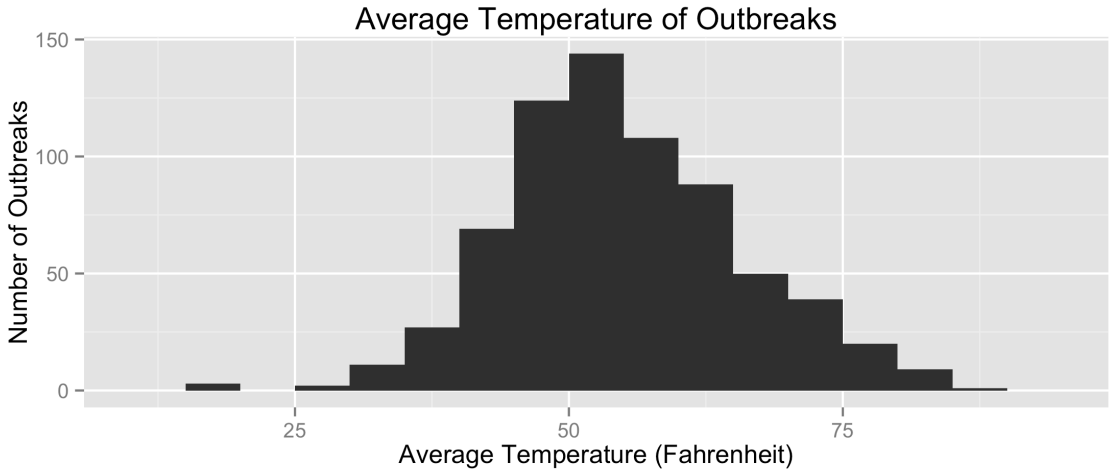
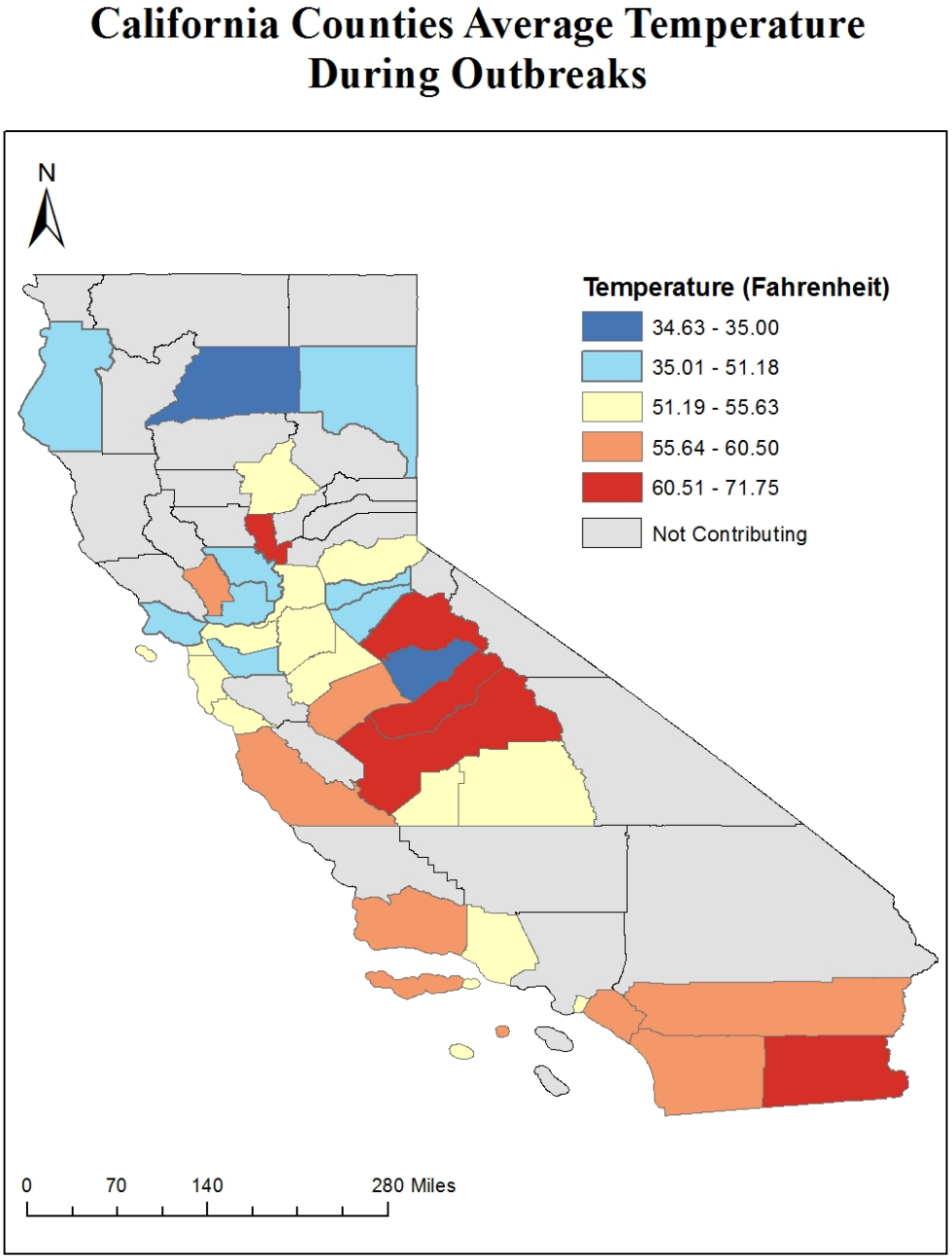


Figure 8 Frequency histogram of average temperature of outbreak

The average temperature during outbreaks was found for each participating county, Map 2 shows the average temperature of outbreak by county. It shows the average temperature is warmer in the south and cooler in the north. The range of average temperatures is evenly dispersed amongst the counties, no one average temperature seems to be dominant. The range of average temperatures in the counties does remain with the 50°F-60°F range.



Map 2 Average temperature of outbreak by county

Rain

The precipitation statistics were created using 695 values. Each outbreak was weighted equally. A q-q plot of the precipitation shows that the data is not normally distributed, Figure 9.

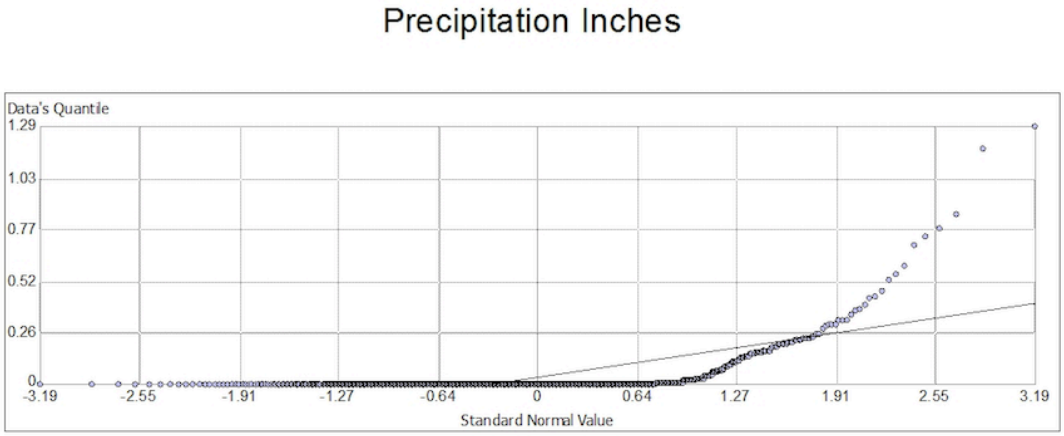


Figure 9 Q-q plot of precipitation during outbreaks

The precipitation frequency histogram, Figure 10, shows that majority of the outbreaks occurred in periods of no to little rain. The mean precipitation during outbreaks was 0.03 inches ± 0.1, the minimum was 0 in. and the maximum was 1.29 in.

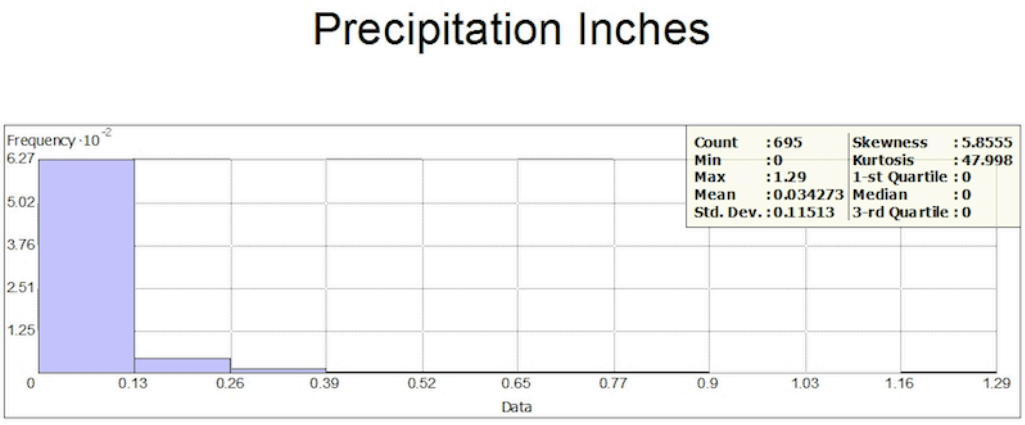


Figure 10 Frequency histogram of precipitation during outbreaks

Wind Speed

There were 611 values for wind speed; this value was not available for all outbreaks. All outbreaks were weighted equally. The values were inputted into a q-q plot, which shows the data is not normally distributed, Figure 11.

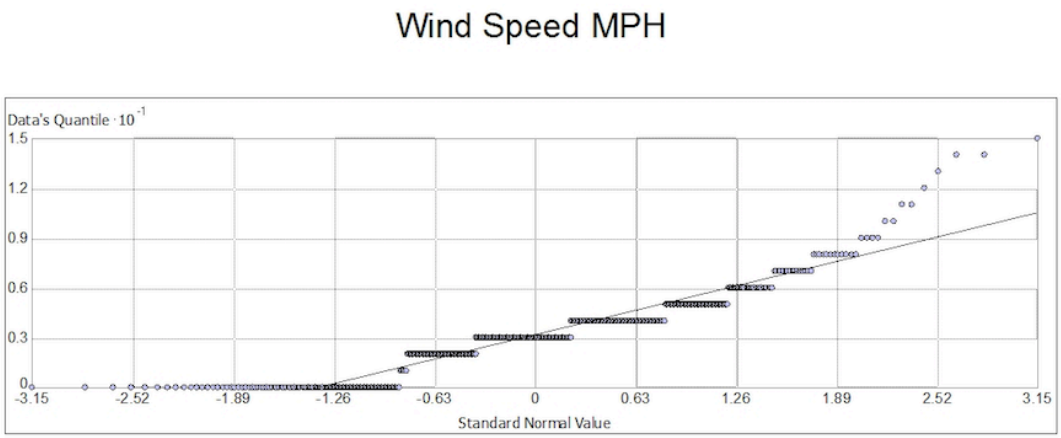


Figure 11 Q-q plot of wind speed during outbreaks

The frequency histogram shows no relationship between wind speed and outbreak occurrence, Figure 12. The average wind speed was 3.13±0.18 mph, minimum wind speed was 0mph and the maximum was 15mph.

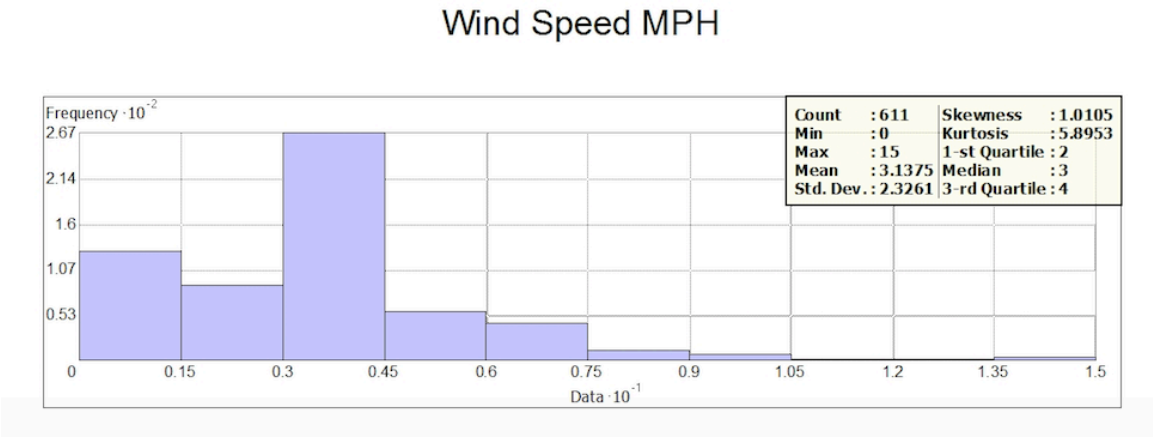


Figure 12 Frequency histogram of wind speed during outbreaks

Humidity

There were 622 values for average humidity; this value was not available for all outbreaks. All outbreaks were weighted equally. The values were inputted into a q-q plot, Figure 13. The results show that there may be a correlation between average humidity and outbreak occurrence.

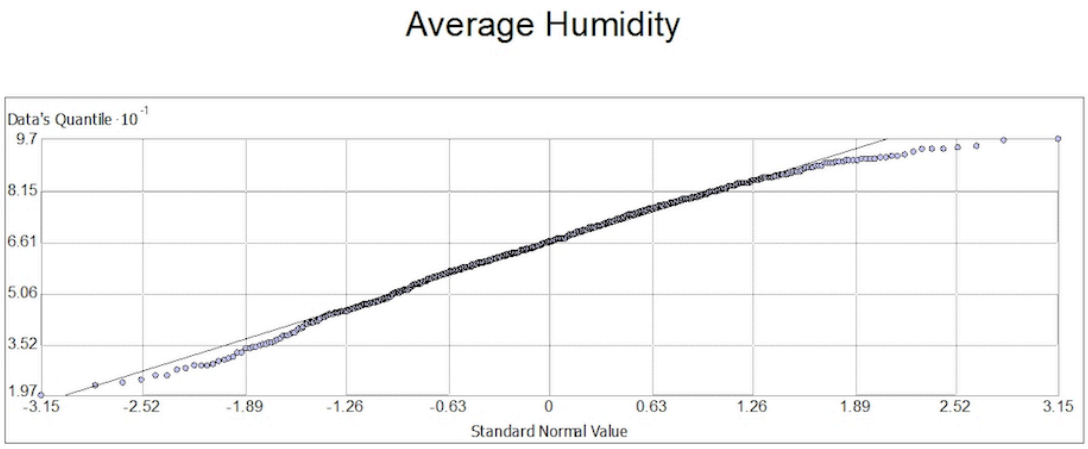


Figure 13 Q-q plot of average humidity during outbreaks

The frequency histogram, Figure 14, also shows a skewed normal distribution. The mean average humidity of outbreaks is 65.23% ±1.19, the maximum is 97% and the minimum is 19.7%.

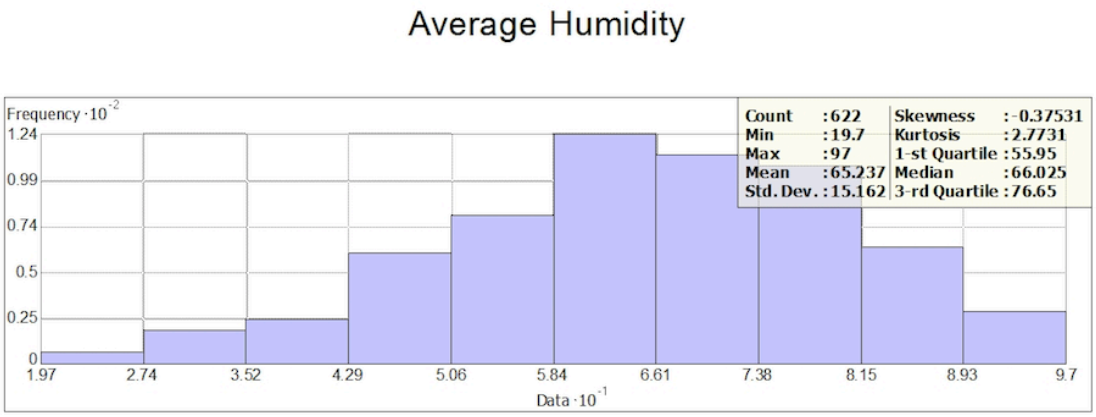


Figure 14 Frequency histogram of average humidity during outbreaks

Looking at a histogram of average humidity, Figure 15 with a binwidth of ten, it is apparent that there is a large range of values that outbreaks occur in. It also shows a skewed normal distribution. A histogram of maximum humidity, Figure 16 with a binwidth of five, shows a similarly right skewed normal distribution. However, when looking at minimum humidity, Figure 17 with a binwidth of five, there is no apparent correlation with outbreaks.

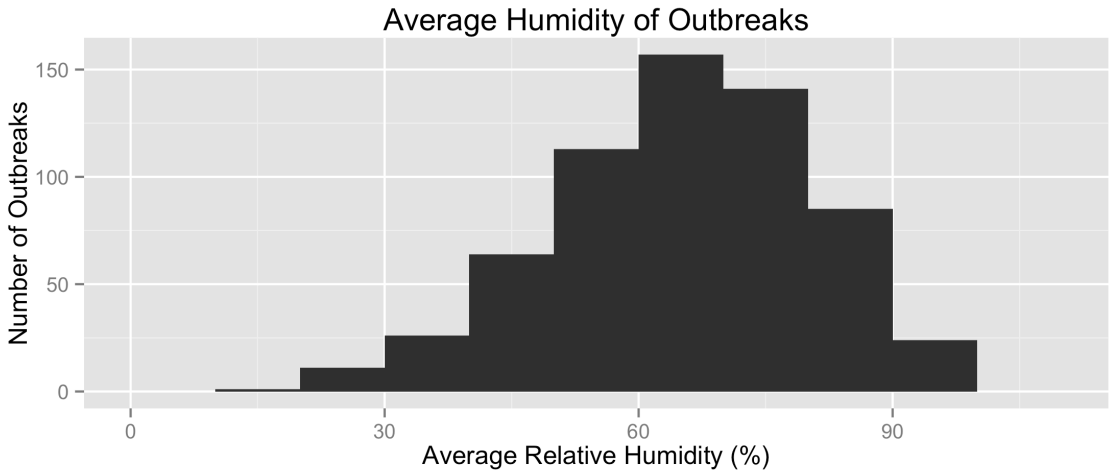


Figure 15 Histogram of average humidity during outbreaks

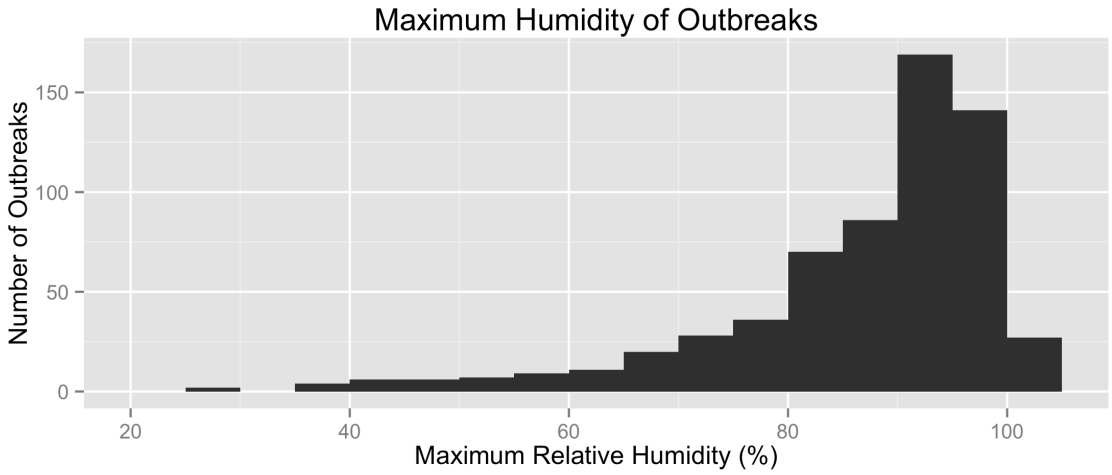


Figure 16 Histogram of maximum humidity during outbreaks

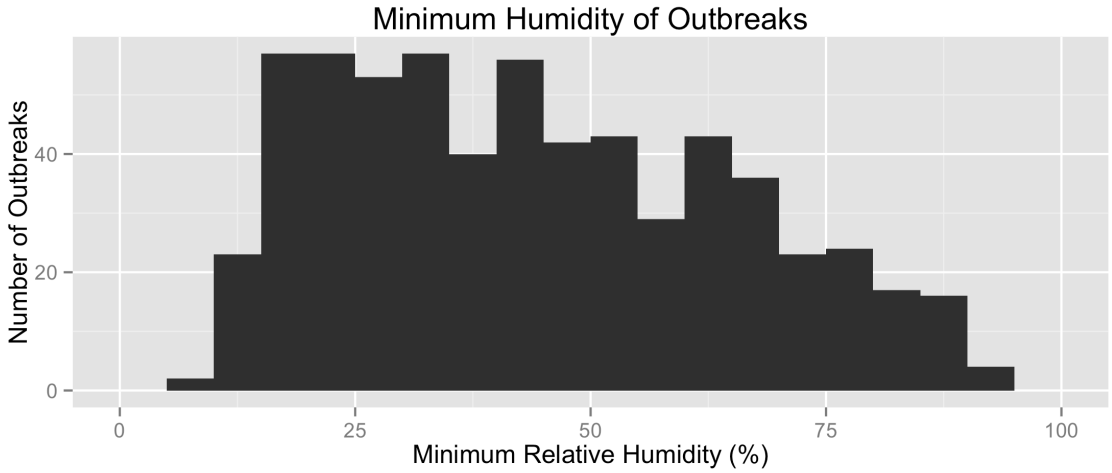


Figure 17 Histogram of minimum humidity during outbreaks

Geospatial Patterns

Mapping the spread

To find geospatial patterns the kriging method was used. First the outbreak data was analyzed to find the peak months of the norovirus season. Figure 18 shows the count of outbreaks occurring in a month. Looking over the multiple norovirus seasons the outbreak typically occurs during October through March, peaking in January. Peaks during January were also seen in a review of norovirus data uploaded to CaliciNet. (Vega et al., 2011)

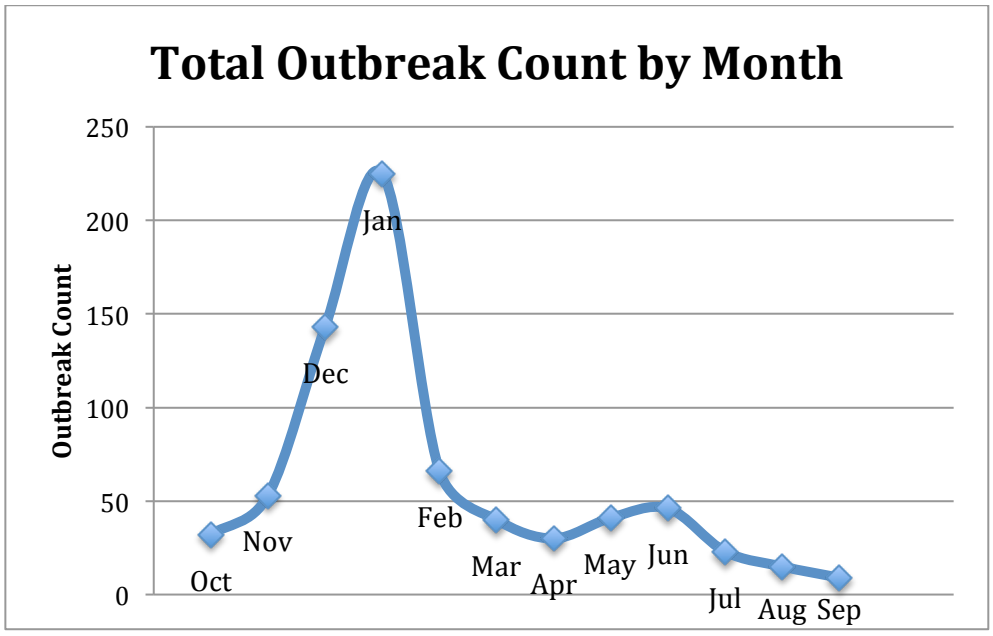


Figure 18 Outbreak count by month

Using this information outbreaks occurring during this time period had their outbreak date transformed into week of the year. The week variable was then transformed into an outbreak week variable. The first week of October became the first week of the outbreak season, week 1, and the last week of March became the last week, week 26.

The kriging tool was used on all seasons of outbreak data available as well as using all of the seasons combined. This tool creates a raster that depicts an estimated surface from points. The magnitude field was populated with the outbreak week variable. An ordinary kriging method was used and a spherical semivariogram model. The output raster will then depict areas in California where different outbreak weeks commonly occurred. This tool was chosen because literature and experts in the field suggest there is a directional bias to norovirus outbreaks during the season. (How Kriging Works, 2012)

The outbreaks were all treated equally and were not weighted for number of lab confirmed cases. Total there were 567 data points with outbreak dates. For the different seasons there were: 2013-2014 83 points, 2012-2013 117 points, 2011-2012 123 points, 2010-2011 23 points, 2009-2010 20 points, 2008-2009 89 points, 2007-2008 42 points, 2006-2007 52 points.

Some of the seasons that had fewer points produced maps that showed little to no spatial patterns. These were seasons: 2006-2007, 2009-2010, and 2010-2011. 2013-2014 produced a map that had no discernible spatial pattern. 2007-2008 produced a map that showed a spatial trend of outbreaks beginning in the north and spreading south over time. This season had few data points, which could allow for skewing. Figure 19 shows outbreak count by month for the entire data set, this shows the years that had fewer outbreaks.

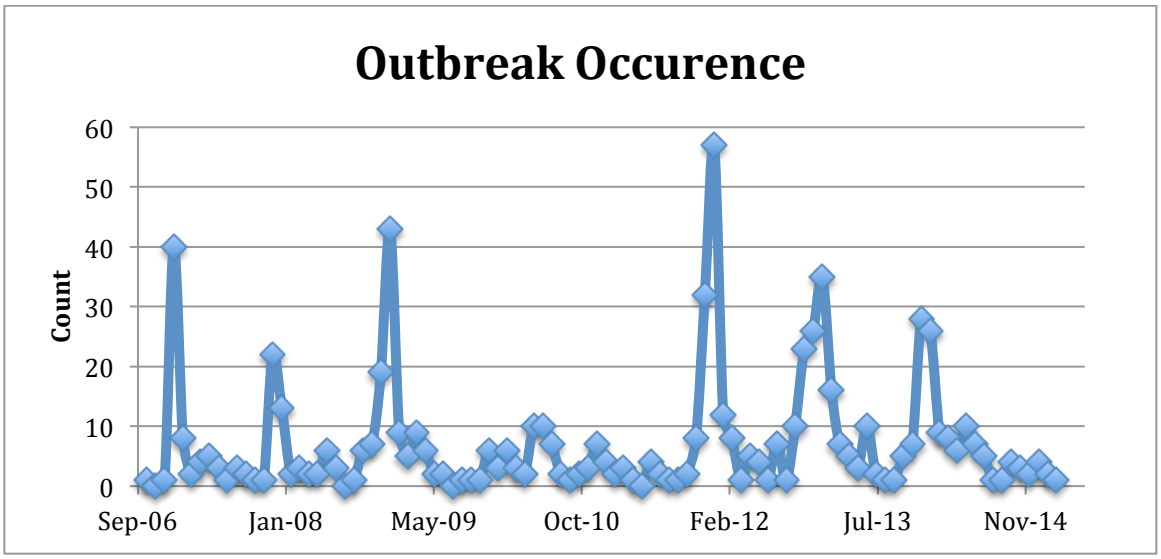
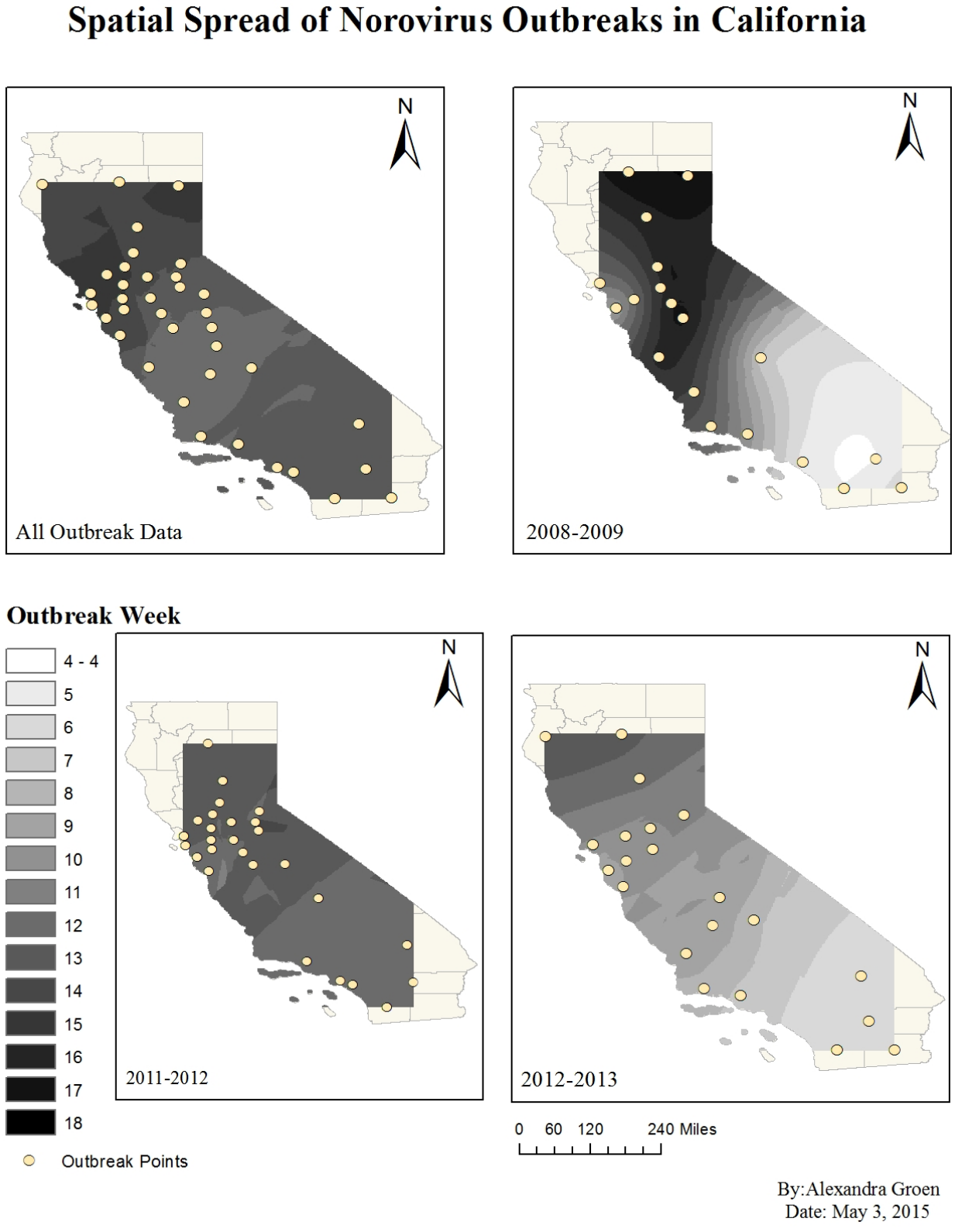


Figure 19 Outbreak occurrence for all years

Several of the maps showed a south to north spread of the virus. 2008-2009 and 2012-2013 had the strongest spatial patterns; both of the seasons were among the seasons with the most outbreaks. 2011-2012 also showed a spread of the virus from south to north but was not as definitive as the other seasons. Using all of the outbreak data to create a map did show outbreaks later in the season clustered in the north, but the trend was not purely south to north. **Error! Reference source not found.** depicts the results for all outbreaks combined and the seasons 2008-2009, 2011-2012, and 2012-2013.



Map 3 Spread of norovirus outbreaks for multiple seasons using the kriging method

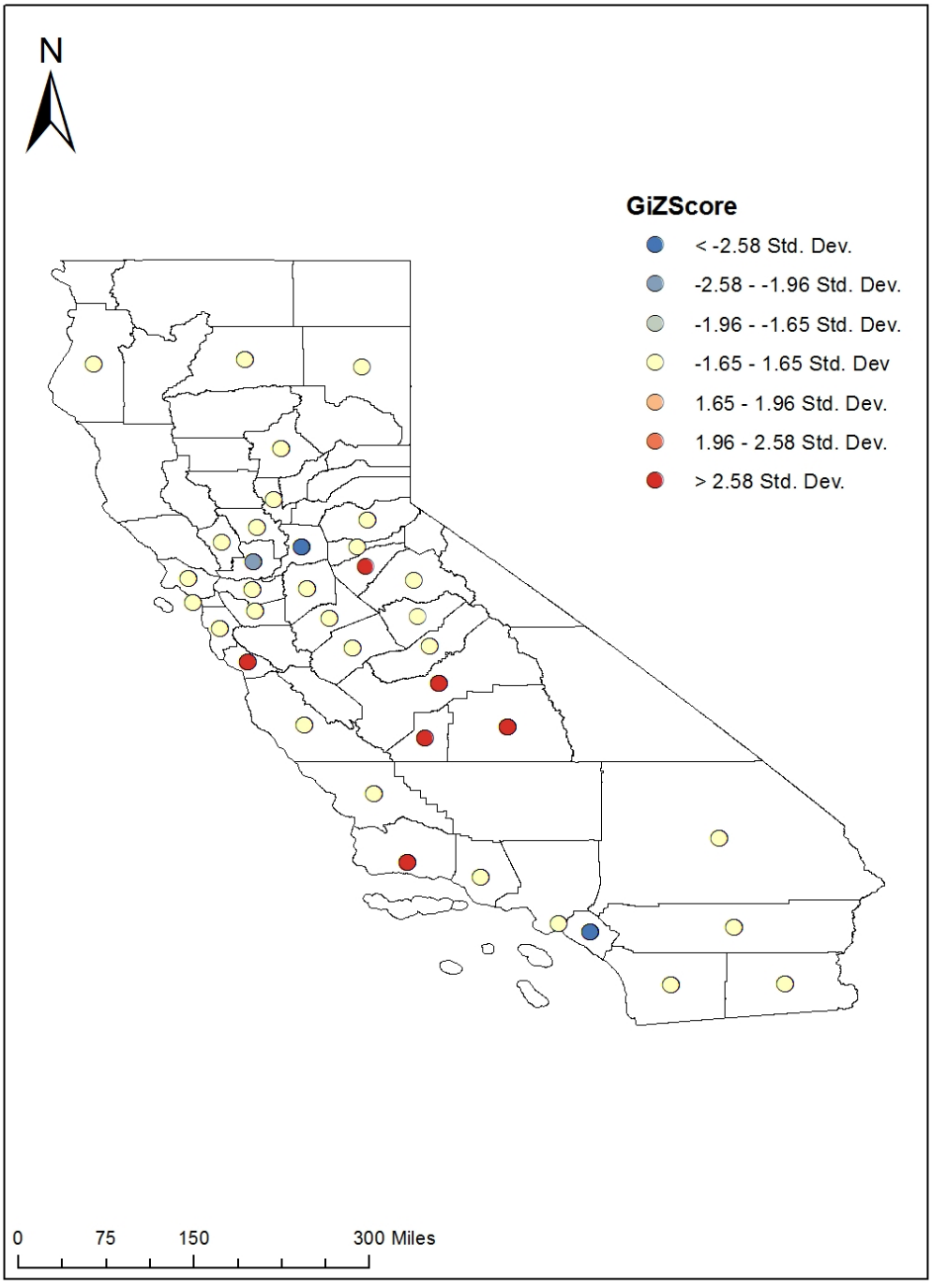
A mean center analysis was also run on outbreaks by month to verify a south to north spread. This analysis takes all outbreak points for each month in the norovirus season and returns a mean center of the points. The results showed the outbreak beginning in the south and the successive months moving further north. The only month that didn't follow this trend was March. This month is the last in the season and has fewer data points than most of the other months.

Hot spot analysis

A hot spot analysis was used to find areas where outbreaks with high number of confirmed illnesses are clustering. This analysis returns a new point feature classified by z-scores that indicate areas where high or low values cluster. Number of lab confirmed cases were used for the input field. (Hot Spot Analysis, 2013)

All outbreak points where used, the number of lab confirmed cases for outbreaks was used as the input field. The output from this will show areas where large outbreaks cluster or where there is clustering of smaller outbreaks. The red and orange points designate "hot spots", areas where there are statistically more large outbreaks, and the blues indicate "cold spots" areas where there are statistically less large outbreaks. Map 4 shows the hot spot analysis results for outbreak data from all seasons.

Hot Spot Analysis of Outbreaks in California

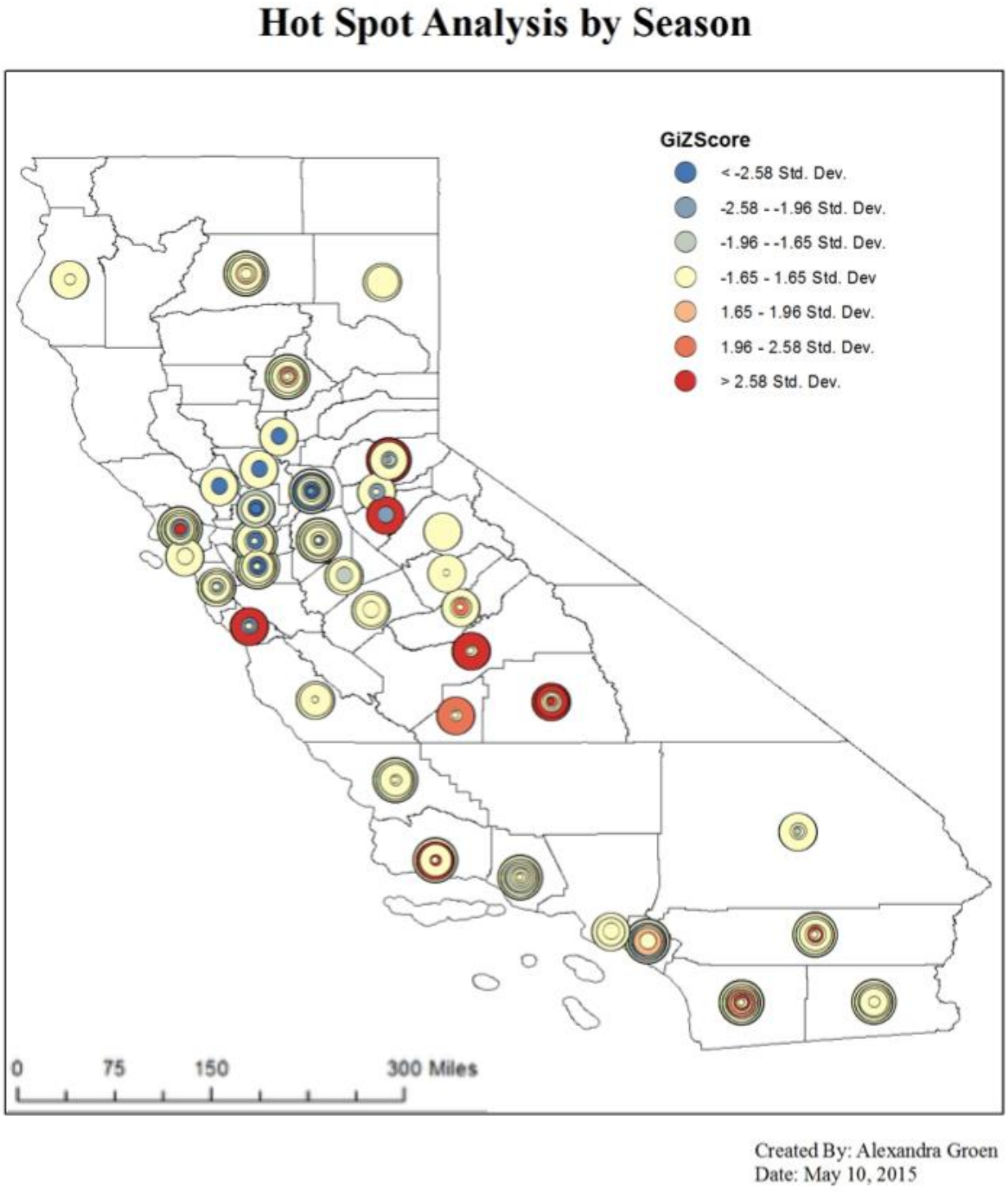


Created by: Alexandra Groen
Date: May 6, 2015

Map 4 Hot spot analysis of all outbreaks

The six hot spots were in counties; Calaveras, Santa Cruz, Fresno, Kings, Tulare, and Santa Barbara. Calaveras, Santa Cruz, and Kings County had ten or less outbreaks. However, their outbreaks had a high lab confirmed number. This caused them to be considered hot spots. The other three counties Fresno, Tulare, and Santa Barbara had twenty or more outbreaks, which had higher lab confirmed counts. They should be potentially considered to be areas where outbreaks infect more people than average. The cold spots were seen in Sacramento, Solano, and Orange County. Solano County had 19 outbreaks, whereas Sacramento and Orange County had seventy and seventy-one outbreaks respectively. These two counties could be considered areas with high numbers of outbreaks that do not infect a large amount of hosts.

To look for skewing of results all outbreak seasons were run independently to find their hot spots. The results were overlaid using graduated symbols Map 5 shows the results. San Diego, Tulare, and Santa Barbara County came up as hot spots in more than one season. Alameda, Sacramento, and Solano County came up as cold spots for more than one season. Some counties were considered hot and cold for different seasons. The results did show an interesting pattern of more cold spots being located in the north and more hot spots in the south.

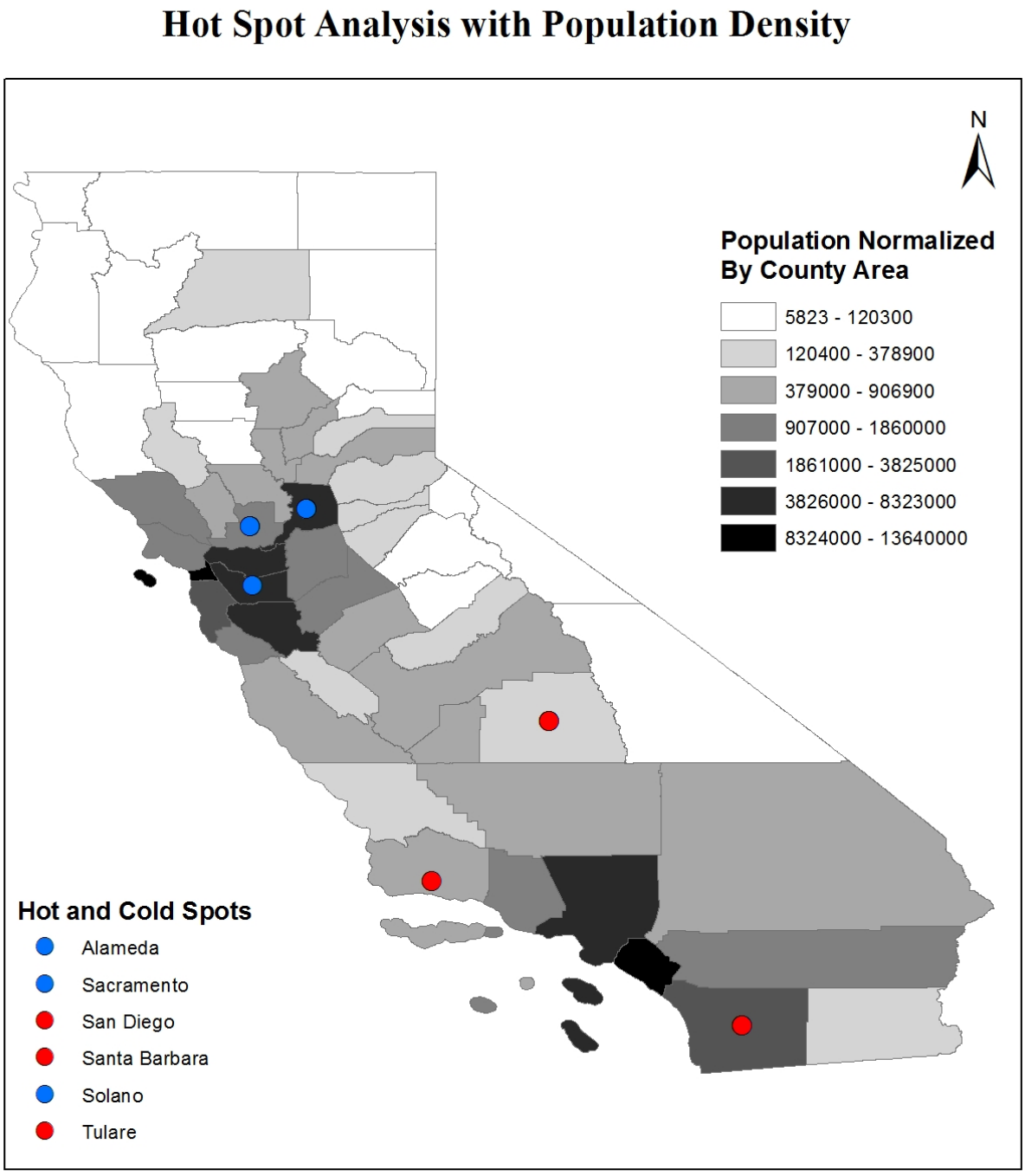


Map 5 Hot spot analysis of outbreaks by season

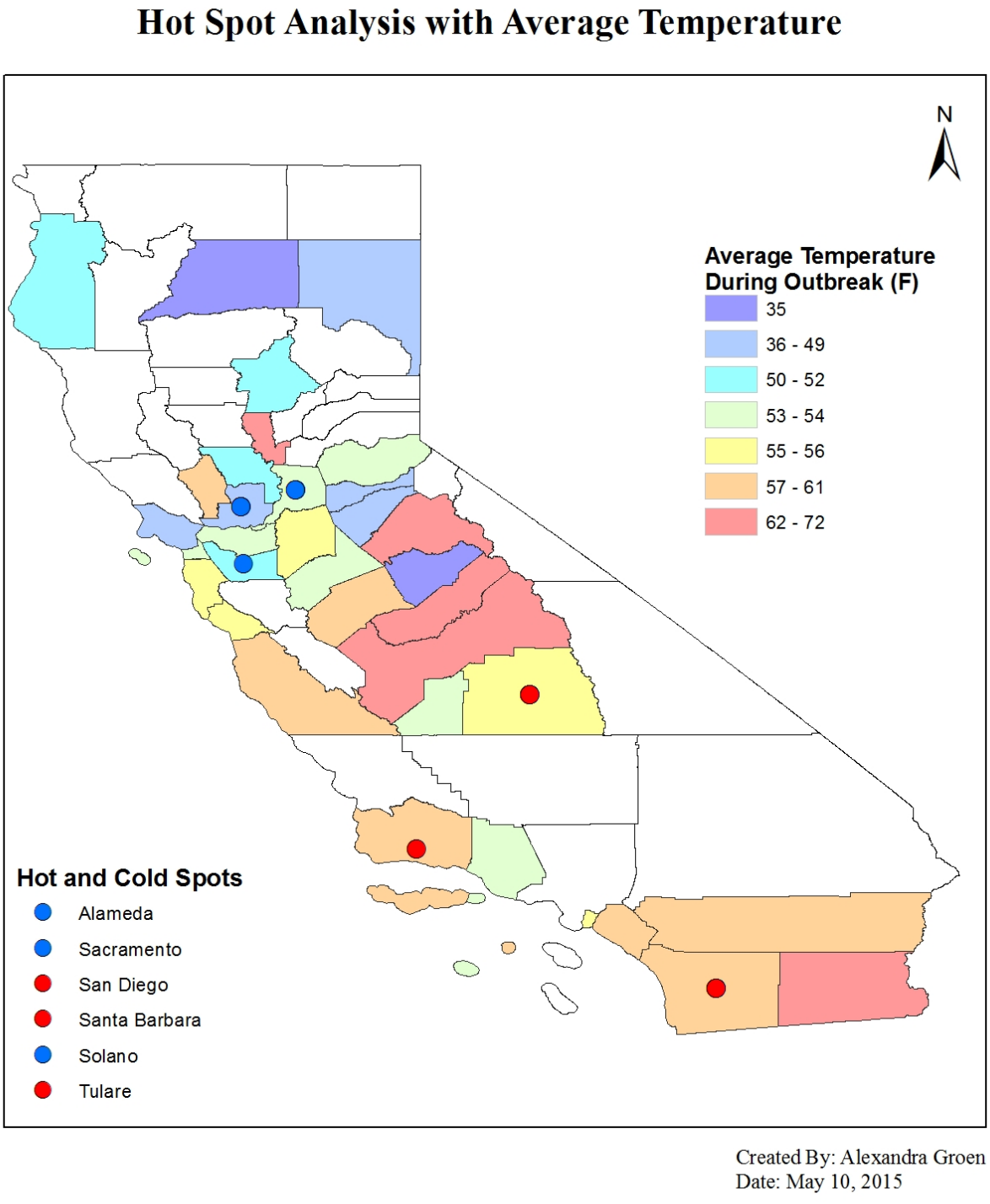
To further investigate the six counties that were hot or cold spots over multiple seasons a chloropleth map of different variables was added to look for relationships. The two chloropleths used were population density, Map 6, and average temperature during outbreak by county, Map 7.

Population density was chosen since one of the main routes of transmission for the disease is person-to-person. If there is a denser population it could facilitate the spread of the virus and cause more outbreaks. Looking at Map 6 no obvious patterns emerge.

Average temperature during outbreak by county was chosen as a layer to further investigate the relationship between temperature and norovirus. Again looking at Map 7 no obvious relationship is seen. It does appear that cold spots are in areas with cooler average temperatures during outbreak and hot spots in warmer average temperatures but this may simply be a function of local climates.



Map 6 Hot spot analysis with population density



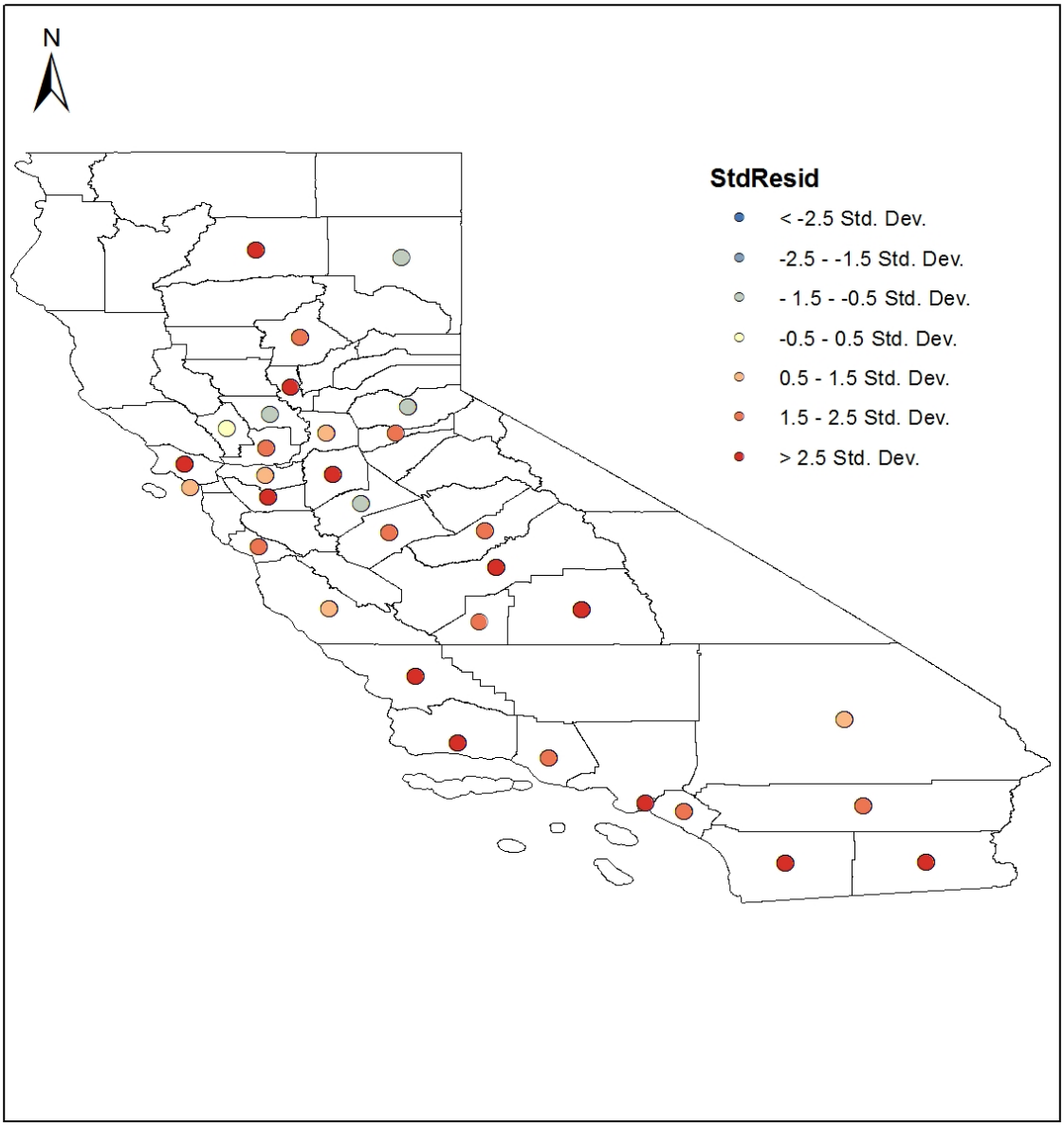
Map 7 Hot spot analysis with average temperature during outbreak

Ordinary Least Squares

Ordinary least squares is a modeling tool in ArcGIS to find correlations between explanatory variables and a dependent value. It returns an output classified by standard residuals. For this model our dependent value was number of lab confirmed cases for an outbreak. The explanatory variables were average temperature, average relative humidity, wind speed, and precipitation.

The model was built to show if there are correlations between large outbreaks and environmental variables. The model output had a low r^2 value, 0.0527. Map 8 shows the results from the ordinary least squares analysis; most of the values have a high standard residual indicating a poor fit.

Ordinary Least Squares Analysis with Environmental Variables



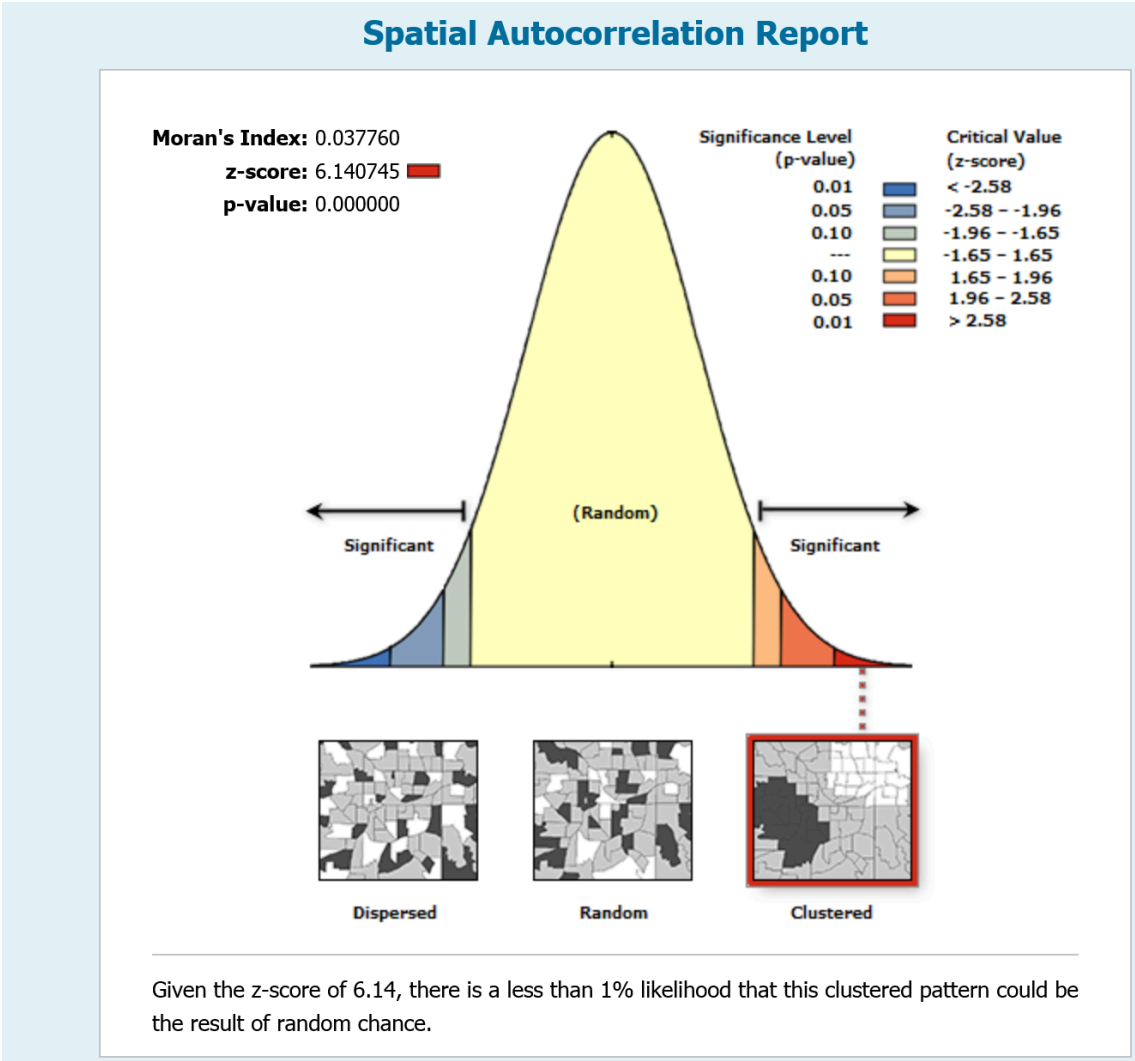
Created By: Alexandra Groen
Date May 10, 2015

Map 8 Ordinary least squares with climate variables for outbreaks

Table 2 Results from ordinary least squares of environmental variables during outbreaks

Variable	Coefficient	Standard Error	T_stat	Probability
Precipitation	-0.428321	0.757867	-0.565166	0.572182
Average Temperature	0.000988	0.007754	0.127443	0.89862
Wind Speed	-0.1868	0.034219	-5.458875	0
Average Relative Humidity	0.000478	0.005774	0.082866	0.933971

Running the spatial autocorrelation tool on regression residuals shows clustering of high variables. This indicates that the model is misspecified, meaning that key variables are missing. For this model it means that we are missing variables that explain clustering of large outbreaks in California. (Interpreting OLS Results, 2013)



Discussion

Environmental variables

Two of the four environmental variables studied showed a possible correlation with outbreak occurrence. As previously discussed there were challenges in sourcing weather data. The analysis in this study could be strengthened considerably by having more precise climate data for where the outbreak originated.

The weather data was also sourced from irrigation stations and airports. This means that the weather stations used were in vastly different environmental settings and can mean variations in the data. Humidity in particular can be skewed

by the differences in stations, since an irrigation station is in agricultural areas and airport weather stations are typically on the tarmac.

Temperature

Overall temperature had a strong correlation with outbreak occurrence. The results show outbreaks typically occur in cooler temperatures, in the 50°F range. The wide range of temperatures supports what is known about how persistent norovirus is. There needs to be more research to find if temperature is actually influencing norovirus outbreaks. The relationship we see may simply be a function of norovirus occurring in the winter months, which have colder temperatures.

One way to begin to see if temperature is influencing norovirus outbreaks is to research if outbreaks are occurring in peak cold times for the county the outbreak originated in. If a pattern is seen it could help to support a theory that temperature influences outbreak occurrence.

Rain

The only pattern seen with norovirus outbreaks and rainfall is that most of the outbreaks occurred when there was little to no rain. This may be a function of the drought that California is currently experiencing. Another study in England looked at rainfall and outbreak and occurrence and found no relationship. Their study was also limited in the precision of their rainfall data. (Lopman, 2009) To see if norovirus is influenced by precipitation a future study should be done in areas that see more rainfall during the winter season.

Rain may still have an effect on norovirus during storm events. Norovirus can survive in water and excess rain could lead to flood events that could facilitate a spread of the virus in an area. (Rohayem, 2009) Further research is needed to see if more outbreaks are seen after flood events to prove this theory.

Wind Speed

There appears to be no correlation between wind speed and outbreaks. A literature search does not show any other studies linking wind and norovirus occurrence. Further research could be done that uses more accurate wind speed measurements for outbreak location as well as wind direction to confidently ascertain if there is no relationship with outbreak occurrences.

Humidity

Humidity showed a normal distribution for average and maximum humidity, and no discernible correlation with minimum humidity. The values for maximum humidity during outbreaks clustering in the 80% and above range indicate that there could possibly be a relationship between high humidity and outbreak occurrence.

In speaking with Mr. Chao-Yang Pan, at the California Department of public Health, he says that increased moisture can help the virus survive on fomites longer. (Personal communication, March 6, 2015) This can help facilitate the spread of the virus, and potentially mean a correlation between high humidity and outbreak occurrence. However other research done in England shows a correlation between outbreaks and a lower relative humidity. (Lopman et al., 2009)

More studies need to be done with more precise humidity data for outbreak location to study the correlation. Influenza is a virus with many similarities to norovirus and it is hypothesized to have a bimodal relationship with absolute humidity. The virus flourishes in low and high absolute humidity but declines in moderate absolute humidity. (Shaman, Goldstein, & Lipsitch, 2011) This could potentially be the relationship that norovirus has with humidity. More case studies are needed to confirm or refute this.

Geospatial Patterns

Spatial Spread

Using all of the outbreak data points to create a map with kriging may have disguised spatial patterns. This is because looking at all of the seasons individually it is apparent that they don't always begin or end in the same weeks. The map does show some clustering of later weeks in the norovirus season in the north and some earlier weeks clustered in the center and southern parts of California. The map does not show a perfect south to north pattern.

The strongest south to north patterns were seen in 2008-2009, 2011-2012, and 2012-2013. The one season, 2007-2008, that showed a north to south spatial pattern had few outbreak points. This means that it could easily be skewed. The earlier seasons have fewer outbreak points; this could be due to limited data collection in the early years. The other years with few outbreak data points, 2009-2010 and 2010-2011, could have few outbreaks due to the development of herd

immunity to the dominant norovirus strain then the influx of points after those years could be due to a new strain emerging. Looking at Figure 20 we can see that the dominant strain GII.4 Minerva peaked in 2008 and a new strain, GII.4 New Orleans, didn't begin to emerge until 2010.

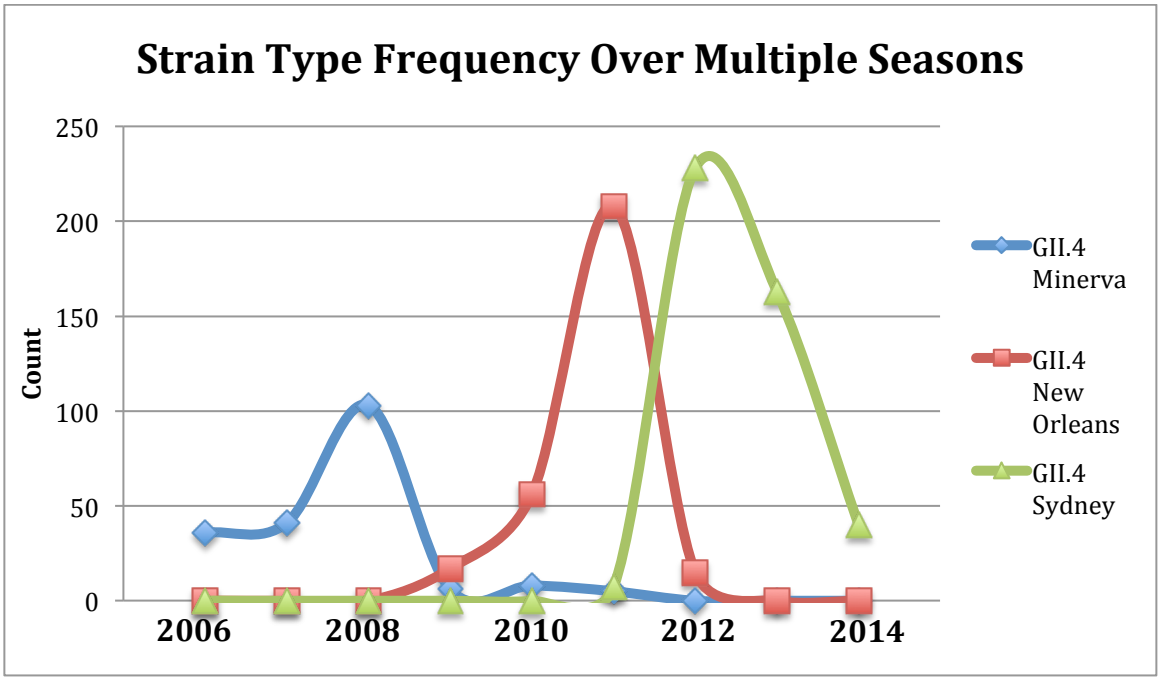


Figure 20 Strain type frequency over multiple seasons

Overall there is evidence supporting a south to north spread of the virus through California. This could be due to the virus following the colder temperatures north. Inaida and the other researchers who looked at the spread of outbreaks in Japan also hypothesized that a south to north spread in a temperate climate could be due to an initial outbreak in the south. (Inaida et al., 2013)

One possibility is that initial outbreaks begin in the south due to Los Angeles being home to the second busiest agricultural port in the United States. In 2011 the ports imports where well over two million metric tons. (Profiles of Top U.S. Agricultural Ports, 2013) This could also correlate with the colder temperatures as agricultural imports are more heavily relied upon in winter months to obtain out of season produce. The virus could be brought in by the crew or by contaminated food. Further studies are needed to see where the first outbreaks during norovirus season

commonly occur then investigate why those areas are seeing outbreaks before other places.

More research is needed on future outbreaks seasons to see if a south to north pattern holds true. This spatial pattern model can be improved upon with better data collection. Such as giving more counties the funds and training needed to be able to test for norovirus so that they can collect data. As well as having a better representation of outbreak location rather than county centroid. A better representation of California will create a better spatial model of outbreak spread.

Further analysis that looks at the rate of spread from south to north would be useful for making basic predictions. If an outbreak is seen in the south and the average rate of spread is known a basic prediction can be calculated that shows when the virus may be seen in the north. This prediction would be rudimentary as it does not include perturbation factors and other variables and would not be exact. However, it could potentially give policy makers and health officials enough advanced warning to begin to implement prevention measures.

Clustering

In reviewing the data in comparison to counties that were classified as hot or cold spots it is evident that better data investigation and classification is needed before this counties can be definitively classified as hot or cold spots. The data was grouped by their outbreak identification numbers to create the lab confirmed category. Some outbreaks in the same county occurred within a day of each other. It is possible these points could have been part of the same outbreak. If this were the case then using the number of confirmed sick for the dependent variable would not be valid.

Ascertaining if patients are sick from the same source and should be classified with the same outbreak identification number can be difficult. Creating different parameters to automatically group individual cases into distinct outbreaks could possibly help to create better and consistent outbreak groups. Different variables such as onset of symptoms and a location buffer could be used to create the parameters needed.

Additionally some counties only have one outbreak recorded. This could be due to clustering but may be indicative of some counties having increased awareness and willingness to test for norovirus. Their needs to be a defined protocol for medical professionals to administer testing for norovirus so that it remain consistent across counties.

Ordinary Least Squares

The results from the ordinary least squares and spatial autocorrelation were not unexpected. It shows that key factors are missing from explaining outbreak size. As literature suggests outbreaks of norovirus are due to a mix of factors, including environmental variables and host behavior. This analysis can be repeated by increasing the number of variables to find what, and what mix of variables, may have a potential impact on outbreak size. Additionally, repeating this analysis on a county scale including weather data for all year may begin to highlight potential correlations with outbreak size and environmental variables. Due to vast range of climates in California patterns may be obscured.

Recommendations

Prevention and Management

Due to little being known about this virus the majority of the prevention and management strategies revolve around the need for more data and further analysis.

Surveillance Networks

In order to increase the ability to test for norovirus and gain valuable data, counties in California that do not have the laboratory setup needed should be given funding to do so. The equipment and training needed is expensive so priority for funding should be given to counties with the highest population. Santa Cruz, Merced, and Butte County have the highest populations of counties who are not able to participate in norovirus testing. They each have over 200,000 people in residence.

Additionally the expansion of states able to participate in CaliciNet will help to capture more outbreak data. The ability to look at outbreaks by state could potentially help look at variables on a wider scale and find overarching patterns. Until there are more laboratories able to test for norovirus it is important to use

syndromic surveillance in conjunction with case studies to look for outbreak patterns.

If norovirus is made to be a reportable disease researchers can be assured that any verified outbreak in the United States is recorded. This can help to identify smaller outbreak clusters and patterns that could have been overlooked.

Future Studies

There is a need for future studies to find correlations with variables that define norovirus occurrences. Studies need to be done on a regional and nation wide scale. Variables and patterns that should be investigated should be sourced from the current literature on norovirus as well as looking at previous studies on viruses that are similar, such as influenza.

Regional studies would allow for a more in depth look at how outbreak occurrence could be connected to environmental variables. As evidenced in this case study, and literature, temperature and humidity are two environmental variables that show potential correlation and should be studied. Additionally, literature suggests that UV exposure can have an impact on host immunity. Decreased UV in the winter could contribute to hosts having lower immunity causing winter seasonality for norovirus. More studies need to be done on these variables in areas where outbreaks occur and precise weather data for outbreak location can be obtained. Studies completed in different states may show similar or different patterns depending on the states climate. It is important to study these variables in a wide variety of climates as patterns found between environmental variables and outbreaks may hold true for different locations with similar climates.

Nation wide studies can help look for overarching patterns in the spread of norovirus and its seasonality. Syndromic surveillance should be used in conjunction with case studies to find variables of importance and spatial patterns. Host behavior should be studied as the literature suggests that is what can contribute to the seasonality of norovirus. Studies done on host behavior could look at travel, agriculture and shellfish imports, time spent indoors. Increased travel could potentially facilitate spread, norovirus can be food borne commonly occurring in leafy greens and shellfish, and increased time indoors could increase person-to-person transmission due to close quarters as well as

increased spread due to studies reporting norovirus can contaminate heating ventilation and cooling systems. Host behavior studies should be based off of current literature about norovirus as well as looking at studies done on influenza, which has similarities in how the virus is transmitted.

Under reporting factors should be found for the United States. This is important so that a predictive model can be properly weighted to find the severity of illness and properly identify spread. A previous study in Germany found that under reporting factors were as high as two to three times. (Bernard et al., 2014)

Conclusion

In conclusion this study found a south to north spread of norovirus throughout its season. The season for California was found to begin in October and end around March. Outbreak occurrence peaked in January. Some correlation was found with temperature and humidity but it requires further investigation.

These conclusions are not enough to begin a predictive model. However, defining the season, outbreak occurrence peak, as well as the pattern of spread can help give some early warning to policy makers and health officials. All health officials should be prepared for an influx of patients with norovirus in January, as that is when the virus peaks. If outbreaks are seen in southern California it would stand within reason that health officials in the north should start preparing for increased number of patients displaying norovirus symptoms.

These results can be strengthened with additional analysis of the data as well as increased surveillance.

Literature Cited

Ahmed, S. M., Hall, A. J., Robinson, A. E., Verhoef, L., Premkumar, P., Parashar, U. D., … Lopman, B. a. (2014). Global prevalence of norovirus in cases of gastroenteritis: A systematic review and meta-analysis. *The Lancet Infectious Diseases*, *14*, 725–730.

Arvelo, W., Hall, A. J., Estevez, A., Lopez, B., Gregoricus, N., Vinjé, J., … Lindblade, K. a. (2013). Diagnostic performance of rectal swab versus bulk stool specimens for the detection of rotavirus and norovirus: Implications for outbreak investigations. *Journal of Clinical Virology*, *58*, 678–682.

Bernard, H., Werber, D., & Höhle, M. (2014). Estimating the under-reporting of norovirus illness in Germany utilizing enhanced awareness of diarrhoea during a large outbreak of Shiga toxin-producing E. coli O104:H4 in 2011 - a time series analysis. *BMC Infectious Diseases*, *14*, 116.

Buehler, J. W., Hopkins, R. S., Overhage, J. M., Sosin, D. M., & Tong, V. (2004). Framework for evaluating public health surveillance systems for early detection of outbreaks: recommendations from the CDC Working Group. *MMWR. Recommendations and Reports : Morbidity and Mortality Weekly Report. Recommendations and Reports / Centers for Disease Control*, *53*, 1–11.

California Counties by Population. (2015). Retrieved Mar 20, 2015, from California Demographics: http://www.california-demographics.com/counties_by_population

Cooper, D. L., Smith, G. E., Regan, M., Large, S., & Groenewegen, P. P. (2008). Tracking the spatial diffusion of influenza and norovirus using telehealth data: a spatiotemporal analysis of syndromic data. *BMC Medicine*, *6*, 16.

Debbink, K., Lindesmith, L. C., Donaldson, E. F., Costantini, V., Beltramello, M., Corti, D., … Baric, R. S. (2013). Emergence of new pandemic gII.4 sydney norovirus strain correlateswith escape from herd immunity. *Journal of Infectious Diseases*, *208*, 1877–1887.

Disease Outbreaks. (2015). Retrieved Mar 20, 2015, from World Health Organization: http://www.who.int/topics/disease_outbreaks/en/

Edelstein, M., Wallensten, A., Zetterqvist, I., & Hulth, A. (2014). Detecting the norovirus season in sweden using search engine data - Meeting the needs of hospital infection control teams. *PLoS ONE*, *9*(6), 1–9.

Glass, R. I., Parashar, U. D., & Estes, M. K. (2009). Norovirus Gastroenteritis. *National Institute of Health Public Access*, *361*(18).

Gormley, M., Templeton, K., Kelly, D., & Hardie, a. (2013). Environmental conditions and the prevalence of norovirus in hospital building drainage system wastewater and airflows. *Building Services Engineering Research and Technology*, *35*, 244–253.

Gunder, L. M., & Dadig, B. (2010). Influenza Overview and Update: 2010-2011. *Clinician Reviews*, *20*(10), 32–37. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=aph&AN=54992866&site=ehost-live

Hall, A. J., Wikswo, M. E., Pringle, K., Gould, L. H., & Parashar, U. D. (2014). Vital signs: foodborne norovirus outbreaks - United States, 2009-2012. *MMWR. Morbidity and Mortality Weekly Report*, *63*(22), 491–5. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/24898166

Henning, K. J. (2004). What is syndromic surveillance? *MMWR. Morbidity and Mortality Weekly Report*, *53 Suppl*, 5–11.

Hewitt, J., Rivera-Aban, M., & Greening, G. E. (2009). Evaluation of murine norovirus as a surrogate for human norovirus and hepatitis A virus in heat inactivation studies. *Journal of Applied Microbiology*, *107*, 65–71.

Hyder, A., Buckeridge, D. L., & Leung, B. (2013). Predictive Validation of an Influenza Spread Model. *PLoS ONE*, *8*(6).

How Kriging Works. (2012, November 8). Retrieved May 5 , 2015, from ArcGIS Resources: http://resources.arcgis.com/en/help/main/10.1/index.html#/How_Kriging_works/009z000000760000 00/

Hot Spot Analysis (Getis-Ord Gi) (Spatial Statistics).* (2013, April 8). Retrieved May 5, 2015, from ArcGIS Resources: http://resources.arcgis.com/en/help/main/10.1/index.html#/Hot_Spot_Analysis_Getis_Ord_Gi/005p000000 10000000/

Inaida, S., Shobugawa, Y., Matsuno, S., Saito, R., & Suzuki, H. (2013). The South to north variation of norovirus epidemics from 2006-07 to 2008-09 in Japan. *PloS One*, *8*(8).

Interpreting OLS Results. (2013, April 18). Retrieved May 5, 2015, from ArcGIS Resources: http://resources.arcgis.com/en/help/main/10.1/index.html#//005p00000030000000

Johnston, C. P., Qiu, H., Ticehurst, J. R., Dickson, C., Rosenbaum, P., Lawson, P., … Perl, T. M. (2007). Outbreak management and implications of a nosocomial norovirus outbreak. *Clinical Infectious Diseases : An Official Publication of the Infectious Diseases Society of America*, *45*, 534–540.

Kirby, A., Iturriza-Gomara, M. (2012) Norovirus diagnostics: options, applications and interpretations. *Expert Review of Anti-infective Therapy* 10,4, 423+.

Lopman, B., Armstrong, B., Atchison, C., & Gray, J. J. (2009). Host, weather and virological factors drive norovirus epidemiology: Time-series analysis of laboratory surveillance data in England and Wales. *PLoS ONE*, *4*(8).

Lopman, B., Gastañaduy, P., Park, G. W., Hall, A. J., Parashar, U. D., & Vinjé, J. (2012). Environmental transmission of norovirus gastroenteritis. *Current Opinion in Virology*, *2*, 96–102.

Modeling Infectious Diseases Fact Sheet. (2014, December 4). Retrieved May 5, 2015, from National Institute of General Medical Sciences: http://www.nigms.nih.gov/Research/specificareas/MIDAS/background/pages/Factsheet.aspx

Mounts, a W., Ando, T., Koopmans, M., Bresee, J. S., Noel, J., & Glass, R. I. (2000). Cold weather seasonality of gastroenteritis associated with Norwalk-like viruses. *The Journal of Infectious Diseases*, *181 Suppl* (Cdc), S284–S287.

Norovirus Diagnostic Methods. (2014, May 12). Retrieved Dec. 4, 2014, from Center for Disease Control and Prevention: http://www.cdc.gov/norovirus/lab-testing/diagnostic.html

Patel, M. M., Hall, A. J., Vinjé, J., & Parashar, U. D. (2009). Noroviruses: A comprehensive review. *Journal of Clinical Virology*, *44*, 1–8.

Prevent the Spread of Norovirus. (2014, Nov. 20). Retrieved Mar. 18, 2015, from Center for Disease Control and Prevention: http://www.cdc.gov/features/norovirus/

Profiles of the Top U.S. Agricultural Ports. (2013, September). Retrieved May 6, 2015, from USDA. http://www.ams.usda.gov/AMSv1.0/getfile?dDocName=STELPRDC5105213

Reporting and Surveillance for Norovirus: Calicinet. (2015, Mar. 15). Retrieved Mar. 18, 2015, from Center for Disease Control and Prevention: http://www.cdc.gov/norovirus/reporting/calicinet/index.html

Reportable Diseases. (2015, Mar. 16). Retrieved Mar. 20, 2015, from National Library of Medicine: http://www.nlm.nih.gov/medlineplus/ency/article/001929.htm

Responding to Norovirus Outbreaks. (2013, Jul 26). Retrieved Apr. 16, 2015, from Center for Disease Control and Prevention: http://www.cdc.gov/norovirus/php/responding.html

Rohayem, J. (2009). Norovirus seasonality and the potential impact of climate change. *Clinical Microbiology and Infection : The Official Publication of the European Society of Clinical Microbiology and Infectious Diseases*, *15*, 524–527.

Samoff, E., DiBiase, L., Fangman, M. T., Fleischauer, A. T., Waller, A. E., & MacDonald, P. D. M. (2013). We can have it all: Improved surveillance outcomes and decreased personnel costs associated with electronic reportable disease surveillance, North Carolina, 2010. *American Journal of Public Health*, *103*(12), 2292–2297.

Sànchez, G., Aznar, R., Martínez, A., & Rodrigo, D. (2011). Inactivation of human and murine norovirus by high-pressure processing. *Foodborne Pathogens and Disease*, *8*(2), 249–253.

Shaman, J., Jeon, C. Y., Giovannucci, E., & Lipsitch, M. (2011). Shortcomings of vitamin D-based model simulations of seasonal influenza. *PLoS ONE*, *6*(6), 1–7.

Shaman, J., Goldstein, E., & Lipsitch, M. (2011). Absolute humidity and pandemic versus epidemic influenza. *American Journal of Epidemiology*, *173*(2), 127–135.

Shapiro, Jason S, MD, MA; Mostashari, Farzad, MD, MSc, MSPH; Hripcsak, George, MD, MS; Soulakis, Nicholas, MS; Kuperman, Gilad, MD, P. (2011). Using Health Information Exchange to Improve Public Health. *American Journal of Public Health*, *101*(4), 616–23.

Specimen Collection. (2013, July 26). Retrieved Apr. 16, 2015, from Center for Disease Control and Prevention: http://www.cdc.gov/norovirus/lab-testing/collection.html

Stals, A., Uyttendaele, M., & Vans Coillie, E. (2013). The need for harmonization in detection of human noroviruses in food. *Journal of AOAC International*, *96*(5), 998–1005.

Van Alphen, L. B., Dorléans, F., Schultz, A. C., Fonager, J., Ethelberg, S., Dalgaard, C., … Lassen, S. G. (2014). The Application of New Molecular Methods in the Investigation of a Waterborne Outbreak of Norovirus in Denmark, 2012. *PLoS ONE*, *9*(9).

Van Beek, J., Ambert-Balay, K., Botteldoorn, N., Eden, J. S., Fonager, J., Hewitt, J., … NoroNet. (2013). Indications for worldwide increased norovirus activity associated with emergence of a new variant of genotype II.4, late 2012. *Euro Surveillance : Bulletin Européen Sur Les Maladies Transmissibles = European Communicable Disease Bulletin*, *18*, 8–9.

Vega, E., Barclay, L., Gregoricus, N., Williams, K., Lee, D., & Vinjé, J. (2011). Novel surveillance network for norovirus gastroenteritis outbreaks, United States. *Emerging Infectious Diseases*, *17*(8), 1389–1395.

Verheyen, J., Timmen-Wego, M., Laudien, R., Boussaad, I., Sen, S., Koc, A., … Pfister, H. (2009). Detection of adenoviruses and rotaviruses in drinking water sources used in rural areas of benin, west africa. *Applied and Environmental Microbiology*, *75*(9), 2798–2801.

Wikswo, M. E., Cortes, J., Hall, A. J., Vaughan, G., Howard, C., Gregoricus, N., & Cramer, E. H. (2011). Disease transmission and passenger behaviors during a high morbidity norovirus outbreak on a cruise ship, January 2009. *Clinical Infectious Diseases*, *52*(January 2009), 1116–1122.